Master Thesis in Statistics and Data Mining

Bayesian poll of polls for multi-party systems

Miriam Hurtado Bodell

Division of Statistics and Machine Learning
Department of Computer and Information Science
Linköping University

Supervisor
Måns Magnusson

Examiner
Anders Nordgaard

Contents

Abstract

This thesis aims to investigate potential poll of polls models for multi-party systems and comparing them with currently used models by finding statistical approaches to evaluating such models, along with analyzing the effects of assumptions regarding data pre-processing and distributions by using dynamic linear models.

Both theoretical and applied results indicate that different strategies of dealing with periodically collected data are of great importance to the results regardless of the model used. The effects of the data pre-processing when creating poll-of-polls models using dynamic linear models has, to our knowledge, never been discussed or studied in the domain and therefore do the results in this thesis show a potential area for future research.

The results in this thesis also indicate that the widely used assumption regarding the independence, using normal approximation allowing for multiple univariate models to be used, between political parties in multi-party system is valid when searching for a poll of polls model measuring vote intention. Dynamic linear models assuming that observed and latent data follows a Dirichlet distribution have so far been unused in the domain, but this novel model outperforms the univariate Gaussian models in five out of eleven of the evaluation measurements and on par in three of them. Using a time variant concentration parameter does not improve the model in a obvious way, but allow for further investigations of the behaviour of the latent state which suggest that Swedish vote intention is more volatile during election campaigns than between elections.

Including house effects seem to be neither beneficial or disadvantageous in poll of polls models in multi-party system using univariate Gaussian models, where using house effects on the variance seem the most appropriate solution.

The attempt of using traditional evaluation measurements, but which are novel within the domain, that distinguishes promising models from a statistical view point and that corresponds with knowledge regarding vote intention from political and behavioural science proved challenging. Where models contradicting expectations based on discoveries concerning Swedish voters are found to be similar to the one generating political polling data.

Acknowledgments

I would like to express my deepest appreciation to my supervisor, Måns Magnusson, who with a never-ending enthusiasm and insight guided me through the process of writing this thesis, working as a constant creative sounding board.

1. Introduction

1.1. Background

In the midst of democratic societies one will find a political process where the people elects representatives to carry out the platforms on which they have been elected. Understanding why people vote in a certain way has been in the interest of scientists from numerous fields during decades. However, elections are held with years in-between but political opinions do not only appear when the election campaigns starts. The opinions about politics and the intention to vote for a certain party could be seen as a latent variable that are measured without bias on election night. Polling houses try to capture the state of the political opinions in-between elections by asking a sample of the population which political party they would vote for if an election were held today.

Understanding the nature of political opinions in a country could be of interest to political parties while trying to find solutions for societal problems whilst also gaining sympathies and streamlining their campaigning and publicity strategies. However, studies with the aim of capturing vote intention in a country are not only useful for political parties but rather anyone with an interest of understanding society. People's opinions in different political issues, mapped to sympathies with different political parties, can be useful to understand the evolvement of several aspects of society, if it reflects on people's behavior. Thus, news events such as civic unrest might also be better understood.

The use of political polls is not a new practice. However, there is still no consensus on how to carry out the aggregation of polls to create a poll of polls model. A brief introduction to the existing research of poll of polls models and modelling of vote intention follows below.

1.2. Previous work

1.2.1. Poll of polls

The use of political polling data in prediction of general elections is a long-standing tradition. However, compared to using polling data in combination with economic variables for analysis and forecasting, the poll of poll approach is less common.

The results of polls can be viewed as realizations of a latent variable corresponding to vote intention amongst a country's population. However, an individual poll is simply a snapshot of the opinions of individuals between elections assumed to correspond with voting intention on the Election Day, and is generally reported with a 95% confidence interval around the presented estimate. If one assumes that the proportions of votes are normally distributed, the precision of an estimate would increase when the sample size increase. Jackman (2005) shows that poll sizes used by polling institutes only yields a small probability of discovering subtle changes in percentage points of the support of a political party. Thus, pooling results from different polls have been proposed to increase the accuracy of the estimates. By pooling results of polls a lower variance, and thus a higher precision, is attained.

Poll of polls have mainly been used in studies of countries with two-party systems, or in countries where the number of established political parties are smaller than in the Swedish political arena, such as the UK. Linzer (2013) uses a poll of polls method to predict the outcome of the US election at a state level, where he comments on the increasing interest for political polling to capture the true level of vote intention and use the today commonly used dynamic linear model to do so. In (Fisher et al., 2011) a similar state-space model was used to capture vote intention in a poll of polls approach to be used for predicting the outcome of the 2010 general elections in the UK. Using a dynamic linear model where the latent state is assumed to be a random walk with drift yields such results that the author expresses clear optimism for future use.

The research conducted in Sweden or other Scandinavian countries is scarce. Walther (2015) uses Swedish polling data to create a poll of polls model, which is then used in combination with economic variables to predict election outcomes. This author raises the difficulties with modeling vote intention in multiparty systems, but evaluates dynamic linear models as having great potential. Stoltenberg (2013) uses different dynamic linear models to create a poll of polls model to predict the general election outcome in Norway, correctly predicting the change of government. An alternative frequentist way of a poll of polls model using Swedish polling data is introduced in (Bergman and Holmquist, 2014) where a compositional loess model is considered a beneficial modeling method since one obtains a single-valued 'residual' measurement of the difference between the poll and the estimations.

The poll of polls approach has been described as having limited prediction abilities if one wishes to predict election outcomes, since the aggregation of polls is often regarded insufficient for this task (Pasek, 2015). Poll of polls are therefore often combined with economic variables when used for prediction, which enables further interpretations of why the party preference is in a certain way. One might also make the claim that political polls consider intention, but if intention does not corresponds completely with the act of voting it can be seen as irrelevant since it only can affect the everyday life of people through the election process.

The view of polls as realizations of a latent variable corresponds with the settings of a dynamic linear model, where the standard is to assume the latent vote share variable to follow a normal distribution (Jackman, 2005; Linzer, 2013; Louwerse, 2015) etc. However, in (Stoltenberg, 2013) the normality assumption is problematized as it is applied in a multi-party setting. The models assuming normality was first adopted in two-party systems where it is based on the belief that the data is generated from a binomial process, which can be approximated with the normal distribution. This is almost guaranteed in a two-party system, but may be more problematic in a multi-party system. The normal approximation runs the risk of working poorly for smaller political parties with vote intention proportions close to zero, especially if the samples are not sufficiently large. Assuming normality of the latent variable does not restrict the vote shares to sum to one either. To avoid the normality assumption Stoltenberg (2013) instead models the observed variable as if it follows a multinomial distribution and the latent a Dirichlet distribution. The result is that the dynamic linear model assuming the latent variable to be Dirichlet distributed produced less volatile predictions, and the author deems it to be a better model choice for Norwegian polling data due to the smoothness of the estimations.

1.2.2. Measuring political opinion

Combining results from polls to create a poll-of polls model is however not a straightforward procedure and several strategies have been proposed. On the blog FiveThirtyEight, the founder and statistician Nate Silver has used models, which give higher weights to newer polls. That is, older polls are considered less reliable and accurate than newer polls when wanting to investigate the current vote intention. However, today there is no consensus regarding what approach yields the best results (Pasek, 2015).

Aside from the issues of how aggregation of polls should be conducted, problems with house effects can arise when modeling poll of polls. ‘House effects’ is the term used for the bias associated with polls produced by polling houses. When conducting surveys there are several sources of error; frame, measurement, non-response and specification (Shirani-Meh et al., 2015). Different houses formulate the questions differently and use different methods for data collection, where many opt for telephone interviews. When conducting telephone interviews persons without phones are excluded, and thus the sample frame and the target population do not match. Further, when polling vote intention the most common question is ‘*What party would you cast your vote for if the election were held today?*’, which could lead to specification errors since wording of the question influences whether you capture the desired measurement. Measurement errors occurs when the survey instruments influence the results. Lastly a non-response error is simply when there is a systematic lack of response that might influence the result. All of these errors may result in house effects.

To correct for the discrepancy between target population and sample frame or for correcting issues with non-response rates in certain groups many polling houses use poststratification. Poststratification is when estimations in the sample are corrected based on strata after the sampling is already conducted, which could be motivated if the average between different groups are known to differ (Lohr, 2010). The variables used in poststratification differ somewhat between houses, and is not necessarily constant for the same house over time. In (Wang et al., 2015) the post-stratification of political polling data was heavily skewed regarding the respondent's age and gender, in relation to the population as a whole. When contacting the polling institutes for further information regarding response rates, most were unwilling to answer questions on the issue. However, one house mentioned non-response rates of around 70 percent, and mentioned that this number probably is similar for other houses. In the report by Oscarsson (2016) on commission of Statistics Sweden the non-response rate of the political polling is 44.2 per cent.

Bergman (2015) finds that house effects indeed are present in Swedish polling data, using a compositional approach, but refrains from naming which houses should be considered more or less trustworthy, but Novus is found to be the closest to the average estimations. There is however no support for that the average estimation is the one closest to the true value of vote intention. Since the house effects where not investigated with an dynamic linear model the results are perhaps not directly comparable to the ones in this thesis. In the study the houses' different methods for data collection, such as random sampling of population registers and the telephone book or self-recruitment, and the varying sample sizes are mentioned as affecting the results. Further, Christensen (2015) publishes on his blog that the Swedish polling institute using web panels continuously shows higher proportion of vote intention for the Swedish Democrats, by using a difference in difference regression, indicating that the method of data collection affects the results of the poll.

Jackman (2005) argues that when pooling polls in the presence of house effects there are no guarantees that the biases will cancel each other out, but instead risks increase the overall bias of the pooled polls. Therefore several attempts to measure and correcting the biases in polls have been conducted, even if research on Swedish polling data is scarce. This is however challenged by Linzer (2013), where the author claims that the bias will cancel out when aggregating over multiple concurrent polls from different houses, finding that the overconfidence in the polls due to potential house effects is minimal.

In (Eady, 2015) the author tries correcting the bias in polls by adding the election outcome as an observation amongst the poll with bias set to zero in a dynamic linear model with Gaussian additive house effects on the mean. Walther (2015) mentioned an alternative approach on Swedish polling data modeling house effects using a median house as a anchor, where the other pollsters' house effect is measured as the difference between the median house and the poll of a certain house. This approach was proposed in (Pickup and Johnston, 2007), with the argument that the industry of political polling as a whole should converge to the truth. However, there is an obvious weakness of this bias correction approach since the median house will still be biased if the houses in general are far away from the truth.

1.2.3. Dynamics of political opinion

Oscarsson (2016) finds that Swedish voters are more volatile today than ever before, with 35 percent of the voters switching party between the 2010 and 2014 election and 17 per cent of voters changing their mind regarding what party to vote for some time between the two elections, based on studies of changes in descriptive statistics. Vote intention is also found to be more volatile than the actual votes cast in the election. That is, voters describe a change in their vote intention between elections but tend to vote for the same party as in previous election on Election Day. However, the study concludes that movements of Swedish voters are limited. Further, tactical voting in Sweden has been documented as increasingly common (Oscarsson and Holmberg, 2011). This indicates a discrepancy between political opinion and vote intention. Oscarsson (2016) also finds that Swedish voters are switching between parties that are ideologically close in regards to the traditional left and right scale. The perception of parties' placement on a political right and left wing scale as presented in the report is shown in Figure 1.1 below.

Figure 1.1.: Perceived political left-right scale before the general election 2014, on the scale 0-100 where 0 is the most left and 100 the most right.

In (Oscarsson and Holmberg, 2015) the descriptive statistics show an increasing proportion of voters making their final decision in the last week before the election in 2010 and 2014, which can be a contributing factor to the difficulties capturing the latent vote intention in polls. However, in Oscarsson (2016) it is found that most Swedish elections do not end in a dramatic election campaign, where voters change their votes in such a way that it changes the outcome of the election in a meaningful way. Right wing parties tend to have stronger finishes in the election campaigns than left winged, where the Left Party is the only one to gain voters in the final days of the campaign more often than what they lose voters for different elections. A recurring pattern is that the largest party of the governing coalition loses more sympathizers before an election compared to other parties.

1.3. Objective

The objective of this thesis is to combine statistical methods with political science to deepen the knowledge and understanding of vote intentions in a multi-party context focusing on Sweden, evaluating the accuracy and performance gains with difference models. This will be done using models adopted directly from research conducted in two-party system as well as using multivariate models reflecting the more complex nature of vote intention being a zero-sum game, and evaluating the potential performance gains by more elaborate models.

The main research question is how one would create a poll of polls model for vote intention that works in a multi-party system. This broad question includes answering the sub questions:

i. How should one deal with periodically collected data in a dynamic linear poll of polls model for a multi-party system?

ii. How do assumptions regarding the distribution of vote intention and polling data affect the models?

iii. Which components should be included in a poll of polls model for a multi-party system?

iv. How can such a poll of polls model for a multi-party system be evaluated?

2. Data

2.1. Data sources

The data consists of political polls from 2006 until today, where the polls keep being added to the dataset when new polls are published. The data comes from an open source GitHub repository[1] provided by one of the cofounders of Botten Ada, a project with the purpose of predicting the outcome of Swedish election using a poll of polls model in 2014. The data collected from 2008 and onwards contains less missing values and is therefore considered as being of higher quality by the distributer. Further, parties are only included if their party sympathies proportions are higher than a specific threshold for consecutive polls in most polls. The Swedish Democrats are therefore not included in polls before 2006 and continues to be missing in certain polls until late 2007. To be able to maximize the number of actual results from general elections all polls from September 2006 until March 2016 will be included in the data used in this thesis. This leads to that 777 polls are used in the data. The results from the general election of 2006, 2010 and 2014 is therefore included in the data as polls without bias, since it is reasonable to believe that election results are measurements of vote intention which contains negligible errors. The information regarding election result is collected from the official website of The Election Authority (Valmyndigheten, 2016).

2.2. Raw data

The data set includes the percentage points that the parties have obtained in the polls as well as what house have conducted the survey. There are in total eleven different polling houses represented in the data. The included polling houses are: Demoskop, Inizio, Ipsos, Novus, Statistics Sweden, Sentio, Sifo, Skop, SVT, United Minds and YouGov. This information will be used for the investigating the presence of house effects.

Information regarding the sample sizes of each poll is included in the data where the smallest reported sample size is around 700 while the largest sample size is almost 13000, not taking election results into account. The largest sample size was reported from a poll conducted by SVT Valu, a poll which is conducted in

[1]Found at https://github.com/MansMeg/SwedishPolls

official election premises asking people that are to cast their votes the day of the election and people that have chosen to vote early (SVT, 2014). This is sometimes referred to as an exit poll and only occurs once for each election. Statistics Sweden differs slightly from the other polling houses since it is an administrative authority and therefore conducts their poll by orders of the government twice every year. The other houses are private companies that conduct polls with frequency of their choice, where most of them conduct more polls close to general elections. Statistics Sweden uses a larger sample than the other pollsters, aside from SVT Valu, with around 9000 participants. Statistics Sweden also has a higher response rate than most of the other polling houses, 50 percent. This can be compared with Novus that uses a sample size of around 4000, with a self-reported response rate of 45-50 percent and Ipsos with a response rate around 25-30 percent in their sample of 1000-1200 respondents. Demoskop on the other hand have difficulties reporting the non-response rate since not all of the telephone numbers included in the sample belong to people or companies that are part of the target population. However, of the people contacted around half are not interested in being a part of the survey.

Sentio, Inizio and YouGov conduct web panels while the others, except for SVT Valu as described above, conduct telephone interviews. However, Sentio also used telephone interviews as their data collection method between 2005 and 2009, switching to a mixture of telephone interviews and web panels between 2009 and 2011, and have since then only used web panels. Sentio uses quotas based on gender, age and region when constructing the sample and report that they use a service where they obtain 1000 responses in each survey. Since the samples from self-recruited web panels are not simple random samples the non-response rate is less meaningful. The polling houses also conduct different weighting of the responses. Statistics Sweden uses the most weighting variables, namely; gender, age, education, country of birth and geographical region. Demoskop, Ipsos, Novus and TNS Sifo uses weights for sex, age and result from the latest election. Sentio only uses the results from the last general election for weighting.

The data also contains information about when the polls have been conducted, both when the data was collected as well as when the poll was published. This information could be used to pre-process the data deciding upon in which time point a poll should enter the model. All polling houses that have answered questions regarding their methodology have confirmed that the responses are equally divided over the collecting period.

Table 2.1 is constructed from the different polling houses webpages where some of the methodology is explained, as well as from mail correspondence with representatives from the different houses. The methodology summary is also composed of information from a summary of an assignment conducted by students at Gothenburg University in the autumn of 2014, which was later complemented and published by Sundell (2015) in the blog Politologerna. The amount of missing values in the table is due to the lack of responses in my own mail correspondence with the houses.

Table 2.1.: Summary of methodology for the polling houses, '-' indicates missing information.

House	Data collection method	n	Response rate	Post-stratification	Methodology change
Demoskop	Telephone interviews	-	30%	Sex, age, education, last election, size	Use more mobile telephone number instead of only landline, previously used 100% landline phone number and today around 40%
Inizio	Self-recruited web panel	2000	-	-	-
Ipsos	Telephone interviews	1000-1200	25-30%	Sex, age, education, last election	-
Novus	Telephone interviews	1000	45-50%	Sex, age, education, last election	-
Sentio	Self-recruited web panel	1000	-	Last election	Oct 2005-Jan 2009 only telephone interviews Feb 2009- Dec 2011 mixture of telephone interviews and web panels Jan 2012 only web panels
TNS Sifo	Telephone interviews	-	-	Sex, age, education, last election	-
Statistics Sweden	Telephone interviews and web panels	9000	50%	Sex, age, education, country of birth, geographical region	Previously only used telephone interviews
SVT Valu	'Face-to-face' interviews	13000	-	-	-
United Minds	Self-recruited web panels	-	-	Last election, internet habits, sex, age	Variables for weighting have increased, previously only used last election
YouGov	Self-recruited web panels	1500	-	Age, sex, geographical region	-

The original data is plotted below, where each poll is colored according to the house conducting the poll, with the election result included as a house called ‘Election’. Overall Demoskop releases polls estimating the vote intention for the Moderates much higher than the rest of the houses until around 2013. One can speculate that this can be attributed to the change of methodology stated by Demoskop through mail correspondence. Sentio on the other hand seems to consistently get results that show a very low support for the Moderates. Demoskop and Skop also have quite high estimations for the Liberals. Further, polls from Sentio and YouGov, both using web panels, are consistently yielding very high results for the Swedish Democrats compared to the other houses, especially obvious towards the end of the time series. This simple visual inspection of the different polls indicates that there exist house effects in Swedish polling data. Further, there seem to be a general trend in the data that the two largest parties, M and S, have lower variance between polls published in the same period than the smaller political parties.

When studying Figure 2.1 below one might also notice that the variation in the time series is highly heterogeneous. That is, in different time periods the vote intention fluctuates more radically while in other time periods the poll results are smoother.

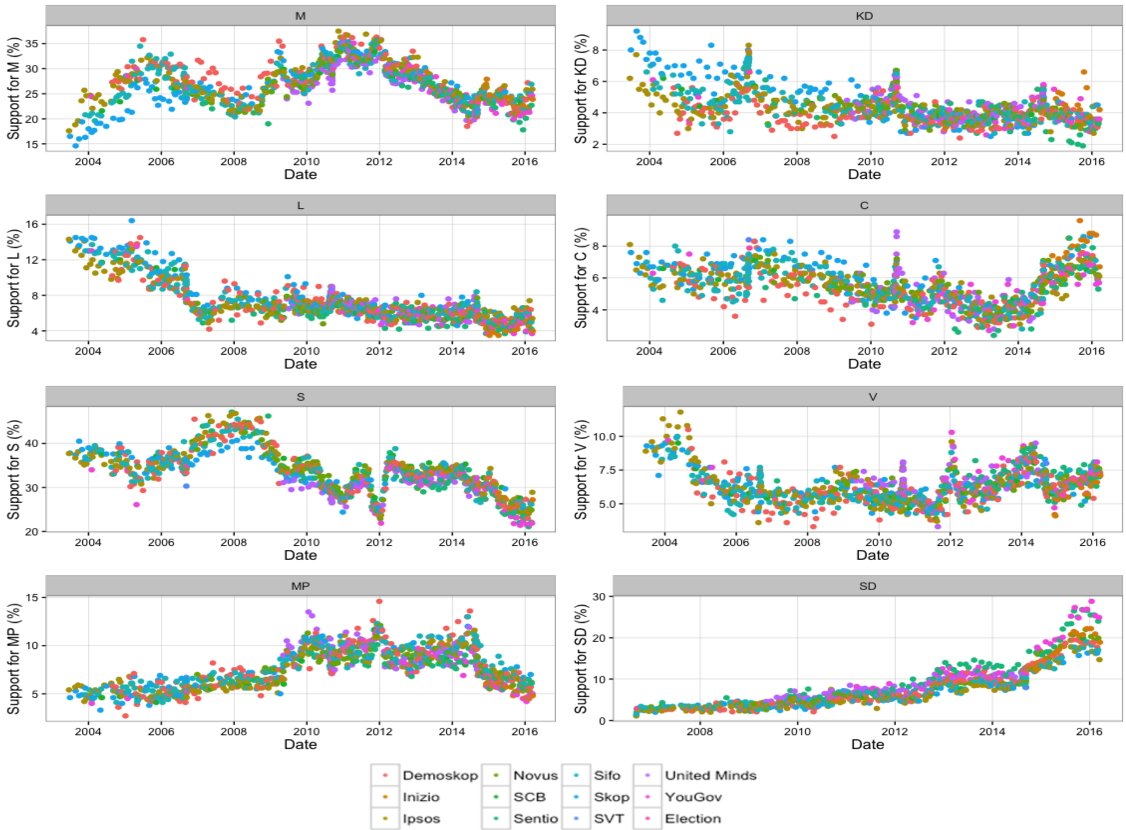

Figure 2.1.: Polling results 2006-2016 coloured by house

3. Methods

3.1. Dynamic linear models

3.1.1. Introduction

Dynamic linear models (DLMs) is mathematically seen as a statistical inverse problem, with the purpose of estimating a latent time series $x_{0:T} = \{x_0, \ldots, x_T\}$ from a noisy observed time series $y_{1:T} = \{y_1, \ldots, y_T\}$. In the Bayesian framework this is equivalent to finding $p(x_{0:T}|y_{1:T})$, which is obtained using Bayes' theorem. In this application the results from political polling is viewed as the observed time series $y_{1:T}$ and the true vote intention seen as the latent time series $x_{0:T}$.

$$p(x_{0:T} \,\big|\, y_{1:T}) = \frac{p(x_{0:T})p(y_{1:T} \,\big|\, x_{0:T})}{p(y_{1:T})} \tag{3.1}$$

$p(x_{0:T})$ is the prior distribution of the latent time series and is defined by the dynamic part of the DLM, $p(y_{1:T} \,\big|\, x_{0:T})$ is the likelihood of the observed data conditioned on the latent time series and $p(y_{1:T})$ is the normalization constant.

The DLM uses two modelling assumptions (Särkää, 2013):

1. The hidden states have the Markov property. That is, the latent variables form a Markov sequence where the x_k given x_{k-1} is independent of everything that happened before time $k-1$.
2. The measurements are considered conditionally independent. That is, y_t given x_t is conditionally independent of historical values of other states or measurements. y_t is therefore only dependent on the position x_t at time t.

The DLMs is divided into two parts, the measurement equation and the dynamic equation. The measurement equation reflects how the measurement y_t depends on the latent state x_t, while the dynamic equation describes the behaviour of the latent variable and how it dependent on its previous state. The generic Gaussian setup is described in the equations in 3.2 where F_t and G_t are assumed to be known transition matrices and V_t and W_t are known covariance matrices (Petris et al., 2009):

$$y_t = F_t x_t + v_t, v_t \sim N(0, V_t) \tag{3.2}$$

$$x_t = G_t x_{t-1} + w_t, w_t \sim N(0, W_t)$$

In this thesis all data is assumed to be available and the main objective is to make inference regarding the latent states that have been observed. This is called a smoothing problem, and the main objective is to estimate $x_{0:T}$ based on the entire sample $y_{1:T}$. That is, the problem involves calculating conditional distributions of $x_{0:T}$ given $y_{1:T}$. This backwards calculation of the conditional distributions starts in the so called filtering density, coming from the corresponding filtering problem when data is assumed to arrive sequentially, which can be defined in equation 3.3 (Petris et al., 2009). The final equality in equation 3.3 holds due to the second modelling assumption stated above.

$$p(x_k \big| y_{1:k}) = \frac{p(x_k \big| y_{1:k-1}) f(y_k \big| x_k, y_{1:k-1})}{p(y_k \big| y_{1:k-1})} = \frac{p(x_k \big| y_{1:k-1}) f(y_k \big| x_k)}{p(y_k \big| y_{1:k-1})} \tag{3.3}$$

in which the prior $p(x_k \big| y_{1:k-1})$ is the one step ahead predictive distribution of the latent state at time k given the observed measurements, which is given as the Chapman-Kolmogorov equation (Särkää, 2013). The Chapman-Kolmogorov equation is a way of expressing how the probability of going from state 1 to state k can be found by using the probability of going from state 1 to an intermediate step and from there to state k, by adding all possible intermediate steps. $p(y_k \big| y_{1:k-1})$ is the one step ahead predictive distribution of the observed values and $f(y_k \big| x_k)$ represent the likelihood. The smoothing density $p(x_k \big| y_{1:k})$ of the state given the observed data is obtained by integrating the backwards transition probabilities in time k (Petris et al., 2009):

1) The conditional distribution of the state vector given the observed data has backwards transition probabilities, the probability of a state given the previous state and all available data, is given by using Bayes' theorem

$$p(x_k \big| x_{k+1}, y_{1:T}) = p(x_k \big| x_{k+1}, y_k) = \frac{p(x_k \big| y_k) p(x_{k+1} \big| x_k, y_k)}{p(x_{k+1} \big| y_k)} = \frac{p(x_{k+1} \big| x_k) p(x_k \big| y_k)}{p(x_{k+1} \big| y_k)} \tag{3.4}$$

2) The smoothing density starts in the filtering density $p(x_T \big| y_{1:T})$

$$p(x_k \big| y_{1:T}) = \int p(x_k, x_{k+1} | y_{1:T}) dx_{k+1} = \int p(x_{k+1} | y_{1:T}) p(x_k | x_{k+1}, y_{1:T}) dx_{k+1} \tag{3.5}$$

$$= \int p(x_{k+1} | y_{1:T}) p(x_k | x_{k+1}, y_k) dx_{k+1} = \int p(x_{k+1} | y_{1:k}) \frac{p(x_{k+1} \big| x_k, y_k) p(x_k \big| y_k)}{p(x_{k+1} \big| y_k)} dx_{k+1}$$

$$= p(x_k | y_k) \int p(x_{k+1} | x_k) \frac{p(x_{k+1} | y_{1:k})}{p(x_{k+1} | y_k)} dx_{k+1}$$

In the classical approach of the state dynamic linear model one assumes that $y_{1:T}$ is Gaussian, as well as both the error term of the observed and dynamic equation.

Since all distributions are assumed to be normal, it follows that all conditional joint distributions from these distributions will also be normal (Durbin and Koopman, 2012) . Using Gaussian distributions results in the Kalman filter where the whole state space model with two equations can be expressed probabilistically as seen below (Petris et al., 2009):

$$
\begin{aligned}
&y_t|x_t \sim N(F_t x_t, V_t) \qquad (3.6)\\
&x_t|x_{t-1} \sim N(G_t x_{t-1}, W_t)\\
&x_0 \sim N(m_0, C_0)
\end{aligned}
$$

However, in most applications F_t, G_t, V_t and/or W_t are unkown, or there are some other unkown parameters in the model, which is the case for the models used in this thesis. The model assumptions therefore changes somewhat where assumption 1 and 2 presented above are belived to hold conditionally on the unknown parameter(s). Thus, the problem is not longer centered about solely making inference regarding the latent state, as presented in equation 3.1, but rather to make inference both regarding the unknown parameter(s) and the latent states by calculating the joint posterior distribution seen below, where ψ represent unknown parameter(s) (Petris et al., 2009).

$$p(x_{0:T}, \psi \,|\, y_{1:T}) = p(x_{0:T}|\psi,\, y_{1:T})p(\psi|y_{1:T}) \qquad (3.7)$$

The joint posterior in equation 3.7 above is calculated through Bayes' theorem but often becomes analytically intractable and MCMC sampling methods can be used to simulate draws from the posterior distributions that inference then can be made on. JAGS have been used for the sampling in this thesis, which is explained more in detailed in appendix A1.

When modelling vote intention both the observed and latent time series are assumed to be between 0 and 1, since they are both considered as being proportions. The sum of the parties, along with an 'Other parties' category should sum to 1. However, no restrictions enforcing these assumptions will be used to investigate which models manage to capture these assumptions unaided.

3.1.2. Dynamic linear models for modelling political opinion

The idea behind the dynamic linear model fits well with the task of detecting the latent vote intention through the measurements provided in political polling. The dynamic equation explained in the previous section represents the dynamics of the behaviour of true vote intention in a population. However, the first model explained here is not a dynamic linear model but rather a standard linear model that will work as a baseline model to compare the other models with. Moreover, one of the objectives of the thesis is to investigate different ways of dealing with periodically collected data in a state space model, presented in section 3.3 as three different data pre-processing techniques. For simplicity all the models below are presented in notation corresponding with data pre-processing technique 1.

3.1.2.1. The benchmark model

The most naive way of conducting a poll of polls is by assuming that vote intention is captured completely in the polls. Thus, the benchmark model only consists of one equation and one cannot add components affecting the outcome of the polls. This model is constructed as a replica of a poll of polls model conducted and published once a month by Novus by commission of a newsroom at Sveriges Radio, a non-commercial independent public service radio broadcaster (SR, 2008).

Since this is a poll of polls model getting media coverage today it can work as a benchmark model, and a standard to which the other models in this thesis should be compared.

The data is pre-processed by combining the results from all polls conducted in a given month, letting the average of these polls represent the mean in the normal distribution vote intention for a specific party is believed to follow. If the collection period of a poll extend over multiple months the result of the poll is divided based on the number of collection days in each of the months, assuming that the data is collected equally over the time period. The poll of polls model of which this is a copy only uses the data from four polling houses: Demoskop, Novus, Sifo and Ipsos, and therefore polls produced by other houses will be left out when modelling this benchmark model(Novus, 2016).

The variance of the model is assumed to follow the variance function for the binomial distribution, and represents the standard error of a poll using a random sample, which is calculated as a function of the poll sample size n_t and the proportion y_{kt} of respondents intending to vote for party k at time t (Jackman, 2005). Thus, the variance of the error term for a single poll is calculated by $y_{kt}(1 - y_{kt})/n_t$. The use of this variance of the error term is convention in research of poll of polls, and might be a 'left-over' from the normal approximation of the binomial distribution used in models for two-party systems. The continued use of this variance can also be motivated by the central limit theorem, where the normal approximation is valid if the number of observations is large enough.

Let μ_{kt} represent the monthly average for party k in month t, in which the vote intention proportion for party k in each individual poll i is represented by y_{ik} with sample size n_i. As explained above it is assumed that the data is collected equally over the period, therefore the data pre-processing can be described as seen in equation 3.8 where a month t consists of T days.

$$\mu_{kt} = \frac{\sum_{i \in t} y_{ik} n_i}{\sum_{i \in t} n_i} \tag{3.8}$$

Thus, this benchmark model can be seen as a process around the monthly average with an error term with a small variance, creating a slowly moving process. The benchmark model can be expressed in the probabilistic notation in equation 3.9, where x_{kt} represents the vote intention proportion for party k in month t each year.

$$\text{Benchmark model: } x_{kt} = \mu_{kt} + \epsilon_k, \epsilon_k \sim N(0, \sigma^2_{kt}) \tag{3.9}$$
$$\text{Constant prior: } \sigma^2_{kt} = \frac{\mu_{kt}(1-\mu_{kt})}{\sum_{i \in t} n_i}$$

3.1.2.2. Basic dynamic linear model

The first dynamic linear model used in the thesis follows the traditional setup for poll of polls models, seen in equation 3.10. The measurement equation is linear where the observed value is assumed to be centred around the latent state with a Gaussian error. The variance in the error term is modelled as described for the benchmark model, as the standard error from a specific poll. The dynamic equation is a random walk, around the previous state of the variable. Compared to the classical set up for a dynamic linear model described in the previous section, both F_t and G_t are seen as time invariant and set to 1 while V_t is the constant parameter and W_t is unknown. y_{ki} represents the proportion of vote intention in poll i for a specific party k, and x_{kt} the hidden proportion of true vote intention for party k at time t in equation 3.10 below. The notation y_{kt} will not be used since multiple polls can be conducted at the same time period t. $\sigma^2_{v_{ki}}$represents the variance in the measurement equation and $\sigma^2_{w_k}$the variance in the dynamic equation.

$$\begin{aligned}
&\text{Measurement equation: } && y_{ki} = x_{kt} + v_{ki}, v_{ki} \sim N(0, \sigma^2_{v_{ki}}) \\
&\text{Dynamic equation: } && x_{kt} = x_{kt-1} + w_k, w_k \sim N(0, \sigma^2_{w_k}) \\
&\text{Priors: } && \sigma^2_{w_k} \sim Gamma(1,1) \\
& && x_{k1} \sim Beta(1,1) \\
&\text{Constant parameter: } && \sigma^2_{v_{ki}} = \frac{y_{ki}(1-y_{ki})}{n_i}
\end{aligned} \tag{3.10}$$

The prior for the error variance is set to a $Gamma(1, 1)$ distribution, which is used to indicate little or no previous knowledge regarding the variance of the error term in the dynamic equation. The initial value for the latent variable is a assumed to follow a beta distribution which is bounded between 0 and 1 and therefore will only yield values that are possible given the nature of proportion of vote intention. $Beta(1, 1)$ will be used as the prior since this is an uninformative prior, equivalent to an $Uniform(0, 1)$, and therefore have equal probability mass for the different possible initial proportions of vote intention.

3.1.2.3. Dynamic linear model with time invariant house effects

As explained in previous work (section 1.2.2) there is evidence of house effects in studies of Swedish polling houses and therefore it is interesting to incorporate a component that captures this in a model. This can be done in different ways, depending on ones beliefs that the house effects are multiplicative or additive, and if the house effects should enter the model through the mean of the measurement model or the variance. A priori it seems more appropriate to introduce house effects through the variance since it does not risk creating problems regarding identification of the model as well as it can be interpreted as how the degree of uncertainty in the surveys differs between polling houses.

The probabilistic set-ups of a model including the house effects will be different, and in this thesis I will explore some of them. Comparisons using different assumptions regarding the house effects and how they should be entered in the model have not been investigated in previous research.

Additive house effects on mean Adding a house effect to the mean is equivalent to adding an explanatory variable to the measurement equation. The set-up of a state space model with one explanatory variable can be represented as seen in equation 3.11 below, which in state space notation would be equivalent to $F = \begin{bmatrix} 1 & \delta_{jk} \end{bmatrix}$ where F is time invariant. In this model the house conducting a specific poll is added to the measurement equation as the parameter δ_{jk}, which is time invariant and estimated by the model. That is, the house effect does not depend on time but only which house $j = \{1, \ldots, 12\}$ that conducts the poll and for which party $k = \{1, .., 8\}$. The house effects used in this model reflect how the mean of the measurement equation are affected by which house is conducting the polling. If a house typically yields higher polling results for a specific political party that house will have a positive house effect, while if a house tends to underestimate a political party the house effect will be negative. Since an additive house effect on the mean could be both negative and positive a Gaussian prior of the parameter will be used. The prior is centred around zero with a variance set to 100, indicating that the house effects are allowed to vary greatly but overall is diminishing. The hyperparameter for the variance of the house effects is equal for all houses, reflecting the same amount of uncertainty of the estimate of the house effect parameter for all houses. The value was chosen based on previous works, where it has been found successful to use on Norwegian polling data (Stoltenberg, 2013). Using this vague prior for the house effects indicates little knowledge of the behaviour of the house effects.

$$\begin{aligned}
&\text{Measurement equation: } && y_{ki} = x_{kt} + \delta_{jk} + v_{ki}, v_{ki} \sim N(0, \sigma^2_{v_{ki}}) && (3.11)\\
&\text{Dynamic equation: } && x_{kt} = x_{kt-1} + w_k, w_k \sim N(0, \sigma^2_{w_k})\\
&\text{Priors: } && \sigma^2_{w_k} \sim Gamma(1, 1)\\
& && x_{k1} \sim Beta(1, 1)
\end{aligned}$$

$$\delta_{jk} \sim Normal(0, 100)$$

$$\delta_{j,Election} = 0$$

Constant priors: $$\sigma^2_{v_{ki}} = \frac{y_{ki}(1 - y_{ki})}{n_i}$$

The model is identified around the election when the true election results are added to the data, with a constant prior for the house effect set to 0. However, an additional restriction on the house effects are needed to get a reasonably working model. Namely that the house effect of a specific house should sum to zero for the different political parties and that the house effects for a specific party sums to zero for different houses. This is a very strong assumption that might reduce the sizes as well as the sign of the house effects which should be kept in mind when analysing the results.

Additive house effects on the variance Adding the house effects to the variance will affect the variability of the model by increasing the variance in the measurement equation.

Measurement equation: $$y_{ki} = x_{kt} + v_{ki}, v_{ki} \sim N(0, \sigma^2_{v_{ki}} + \delta_{jk}) \tag{3.12}$$

Dynamic equation: $$x_{kt} = x_{kt-1} + w_k, w_k \sim N(0, \sigma^2_{w_k})$$

Priors: $$\sigma^2_{w_k} \sim Gamma(1, 1)$$

$$x_{k1} \sim Beta(1, 1)$$

$$\delta_{jk} \sim Gamma(1, 1)$$

$$\delta_{j,Election} = 0$$

Constant priors: $$\sigma^2_{v_{ki}} = \frac{y_{ki}(1 - y_{ki})}{n_i}$$

To avoid the risk of obtaining a negative variance the prior for the house effects needs to be semi-infinite $[0, \infty]$, and therefore a $Gamma(1, 1)$ is used. This prior will lead to that the variance in the measurement equation can only increase or be the same, which is an obvious weakness of the model. This prior is however used to avoid obtaining a negative variance, which would have been possible if a prior covering negative values would have been used. A house effect close to zero will indicate that the variance is the same as to be expected in a random sample, and therefore the prior for the house effects for the elections is set to 0. As before the house effects will be different for the different parties for each house.

Multiplicative house effects the variance Using a multiplicative house effect on the variance indicates that the variance of a poll for a specific house may be higher or lower than what one could expect when using a random sample, which is sometimes called a design effect (Fisher et al., 2011). By entering the house effects multiplicatively through the variance one captures the natural interpretation of the parameter as affecting the assumed variance in a poll.

A house that has a high positive house effect will decrease the precision of a poll, and thus increasing the variance of that poll. When the variance is increased it allows for the variability in the model to depend not solely on the estimated value of the latent variable but rather as noise in the measurement. If the house effect is below 1 the variance of the measurement equation will decrease, indicating a higher precision of an observed result.

Once again an uninformative gamma prior is used for the house effects to avoid negative values for the combined variance. The election results will again be assumed to be without bias, and therefore is set to 1 since this will not change anything in the measurement model from the basic dynamic linear model. No assumption regarding the sums of the house effects are made since the identification problem is not an issue using multiplicative house effects on the variance.

$$\begin{aligned}
&\text{Measurement equation: } && y_{ki} = x_{kt} + v_{ki}, v_{ki} \sim N(0, \sigma^2_{v_{ki}}\delta_{jk}) && (3.13)\\
&\text{Dynamic equation: } && x_{kt} = x_{kt-1} + w_k, w_k \sim N(0, \sigma^2_{w_k}) &&\\
&\text{Priors: } && \sigma^2_{w_k} \sim Gamma(1,1) &&\\
& && x_{k1} \sim Beta(1,1) &&\\
& && \delta_{jk} \sim Gamma(1,1) &&\\
& && \delta_{j,Election} = 1 &&\\
&\text{Constant priors: } && \sigma^2_{v_{ki}} = \frac{y_{ki}(1-y_{ki})}{n_i} &&
\end{aligned}$$

3.1.2.4. Dynamic linear model with time variant house effects

The house effects investigated so far have been time invariant, meaning that they do not depend on time and are thus considered to stay the same over the whole time series. However, some of the houses have stated changes in their methodology over time that could have had an effect on the potential bias of the houses. It would not be improbable that the bias associated with a specific house would evolve over time.

The time variant house effect will be added in a way that is found the most promising when using different models for time invariant house effects presented in the previous chapter. The time variant model will therefore have almost the same set up as 3.11,3.12 or 3.13 with the difference that the house effect will be represented by δ_{jkt}

instead of δ_{jk}. The prior value of the house effect will be drawn either as a $N(0, 100)$ or a $Gamma(1, 1)$, depending on if the most appropriate model use house effects on the mean or variance. The house effects thereupon over time will then be drawn from a normal distribution with the house effect estimated at the previous time-point as the mean, and with a variance that is estimated by the model. One can make different assumptions regarding the variance of the random walk house effects. If one believes that the house effects are mainly due to changes in methodology the variance should be large to allow for sudden large steps in the house effects. If the house effects instead are induced by the sensitive nature of political polling questions or by constant problems with non response rates the random walk process would be a slower moving process. Here a $Gamma(1, 1)$ will be used to estimate the variance term, indicating a slowly moving random walk process.

$$\text{Priors:} \qquad \delta_{jkt} \sim Normal(\delta_{jkt-1}, \sigma_\delta^2) \tag{3.14}$$
$$\sigma_\delta^2 \sim Gamma(1, 1)$$

This would indicate a less restrictive model, allowing for the bias associated with the different houses to change over time enabling evaluations of methodology changes regarding data collection.

3.1.2.5. Basic Dirichlet-Dirichlet model

In the previous models normality of the vote intention variable and the measured polling data was assumed, and the models have been univariate treating the different parties independently. However, in a multiparty system the results from polls could not be seen as from a Bernoulli trial, since the question asked in the polls are '*What party would you vote for if the election were held today?*' to which there are multiple answers. Thus, the answers to polls could be considered as generated by a multinomial distribution. However, since the results of the polls are presented in terms of percentages of respondents intending to vote for a specific party the measurement equation could be seen to follow a Dirichlet distribution.

The Dirichlet distribution can be parameterized with a vector of positive real values, one for each possible category $\boldsymbol{p} = \{p_1, \ldots, p_9\}$ and a concentration parameter α, where $\boldsymbol{p}$ sums to 1. The number of categories will be 9, one more than the 8 parties of main interest since there are some people intending to vote for parties that are not represented in parliament. The concentration parameter α controls how centred the distribution is around the mean. A small α will favour extreme values and a high valued α will yield a distribution closer to $\boldsymbol{p}$ (Frigyik et al., 2010). The concentration parameter can therefore be seen as a variance parameter that reflects the movement of the latent variable between time t and time $t + 1$. The size of the concentration parameter can also be seen as the additional information from a new random sample of the same size of the concentration parameter. It is therefore natural to use the actual sample size n_i of the poll as the concentration parameter in the Dirichlet distribution of the measurement equation. This notation differ from how the Dirichlet distribution is presented in most standard text books, e.g. Gelman et al. (2014), where the distribution has only the parameter vector α with one element for each possible category. However, both ways would yield the same distributions due to the restriction of $\boldsymbol{p}$ summing to 1 and that the values of the elements in $\boldsymbol{\alpha}$ are positive reals..

The latent states should reflect the probability of a respondent intending to vote for a specific party, and these probabilities should sum to 1. Therefore is it appropriate to assume that the latent states also follows a Dirichlet distribution. The output of the Dirichlet distribution consists of values between 0 and 1 and $\sum_{i=1}^{9} x_i = 1$, which corresponds well with the nature of vote intention. Since the appropriate size of the concentration parameter is unknown a $Gamma(1, 0.0001)$ prior will be used to estimate it within the model. This prior allows for a broad spectrum of values and is therefore very uninformative. A high value of the concentration parameter indicates that the variance decreases and that the dynamic states are smoother, while a low value of the parameter indicates high volatility of the latent states. Therefore one can speculate that the concentration parameter should be estimated rather high, which further justifies that the mean of the chosen prior is 10000.

The probabilistic set-up of the model can be seen below. The use of a Dirichlet distribution of both the measurement and dynamic equation is a novelty, as well as the estimation of the concentration parameter α. However, one should keep in mind that the concentration parameter of the dynamic equation is assumed to be constant over time, which corresponds to the belief that the dynamics of vote intention is constant. This assumption is not consistent with previous research regarding Swedish vote intention but previously used successfully on Norwegian data (Stoltenberg, 2013). The prior for the first value of the latent states is an uninformative Dirichlet prior, assuming equal probabilities for the categories and the concentration parameter is 1.

$$
\begin{aligned}
&\text{Measurement equation: } && p(\boldsymbol{y}_i) \sim Dirichlet(\boldsymbol{x}_t, n_i) \qquad (3.15)\\
&\text{Dynamic equation: } && p(\boldsymbol{x}_t) \sim Dirichlet(\boldsymbol{x}_{t-1}, \alpha)\\
&\text{Priors: } && \boldsymbol{x}_1 \sim Dirichlet(\frac{1}{9}, ..., \frac{1}{9}, 1)\\
& && \alpha \sim Gamma(1, 0.0001)
\end{aligned}
$$

where the vector notation above is $\boldsymbol{x}_t = \{x_{1t}, ..., x_{9t}\}$, $\boldsymbol{y}_i = \{y_{1i}, ..., y_{9i}\}$.

3.1.2.6. Dirichlet-Dirichlet model with a time variant concentration parameter

As mentioned in section 3.1.2.5 above one drawback with the basic model can be the assumption that the dynamics of the latent variable is constant over time. In previous work regarding the dynamics of political opinions (section 1.2.3) one can read that this is not necessarily a realistic assumption. Therefore this restriction is dropped in an extended Dirichlet-Dirichlet model where the concentration parameter of the latent state is seen as time variant, evolving over time. Since the polling data shows turbulence in certain time periods, visible in Figure 2.1 corresponding to the Juholt scandal for the Social Democratsin in the end of 2011 leading to that the party leader, Håkan Juholt, eventually had to resign due to his unpopularity hurting the party, this might be reflected in the estimation of the concentration parameter. In times of higher volatility the concentration parameter should decrease, meaning that the concentration parameter can be seen as a stability parameter, with high values indicating high stability. The initial value of the concentration parameter will be drawn from the $Gamma(1, 0.0001)$ distribution, which is the same vague prior used in equation 3.15, while the prior for the rest of the values is a normal distribution with the concentration parameter of the previous time point as the mean and with a $Gamma(1, 0.001)$ prior for the variance. This prior of the variance for the concentration parameter is vague since it allows for both very small and large values, enabling both rapid and smooth changes for the concentration parameter.

$$\begin{aligned}
&\text{Measurement equation: } && p(\boldsymbol{y}_i) \sim Dirichlet(\boldsymbol{x}_t, n_i) \\
&\text{Dynamic equation: } && p(\boldsymbol{x}_t) \sim Dirichlet(\boldsymbol{x}_{t-1}, \alpha_t) \\
&\text{Priors: } && \boldsymbol{x}_1 \sim Dirichlet(\frac{1}{9}, ..., \frac{1}{9}, 1) \\
& && \alpha_1 \sim Gamma(1, 0.0001) \\
& && \alpha_t \sim Normal(\alpha_{t-1}, \sigma^2_\alpha) \\
&\text{Hyperparameter priors: } && \sigma^2_\alpha \sim Gamma(1, 0.001)
\end{aligned} \tag{3.16}$$

3.2. Model diagnostics and evaluation

When sampling from the posterior distribution one needs to examine convergence to the target distribution. Studying the produced trace plots for the parameters of interest will do this. Trace plots show the value drawn from the posterior distribution at each iteration, and is therefore a visual inspection of how well the samples explore the posterior distribution. The trace plots should show randomness around the mean of the parameter of interest if the model has converged. If the posterior distribution has multiple peaks the MCMC risks getting stuck at one of the modes.

The evalutaion measured used in this thesis is presented in Table 3.1 below, followed by a more indepth explanations of each measurement.

Table 3.1.: Summary of evaluation measurements.

Evaluation measurement
95% central credible bands
MAD
RMSE
Posterior RMSE
Posterior predictive check

When evaluating a poll of polls model one method has been to investigate if the model has captured the real election results for the different political parties within a certain credible band. In a $100(1-\alpha)\%$ central credible band α is chosen based on how large proportion of the posterior distribution that should be included in the interval. If α is set to 0.05, 95% of the simulated values from the posterior are within in the bands. This is an intuitive evaluation method if one sees the elections as a poll of vote intention without any bias. It is however somewhat limiting since elections are far apart, making evaluation possible only on a few data points. When evaluating the performance of the models of election results from 2010, only data collected before the day of the election 2010 are used. The same is done when predicting the election result 2014. The latent time series are modelled to also cover the Election Day, and one can therefore compare the simulations from the posterior distribution at this time point to the election result. The point estimate used will be the expected value of the posterior at the day of the election. Mean absolute deviation (MAD) and root-mean-squared error (RMSE) between the expected values of the posterior distributions of the Election Day and the election results will be calculated as a way of evaluating the models, which are calculated using formula 3.17 and 3.18 below, where $\hat{x}_i$ is the point estimate for each of the eight parties and x_i is the election result for party i, $i = 1, ..., 8$ and k is the number of parties.

$$MAD = \frac{1}{K}\sum_{i=1}^{k}|x_i - \hat{x}_i| \tag{3.17}$$

$$RMSE = \sqrt{\frac{1}{K}\sum_{i=1}^{k}(\hat{x}_i - x_i)^2} \tag{3.18}$$

To deepen this analysis an additional measurement will be used, which will be called the Posterior RMSE (PRMSE), where each of the simulations from the posterior distribution of the latent state on Election Day will be compared to the true election results. This will capture both an average distance between posterior point estimates and the true value, as well capturing the variance in the posterior. This measurement is formulated below, where the first term in the sum is drawn i from the posterior distribution from the Monte Carlo simulation of the latent variable at the time of the election.

$$PRMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{x}_t^{(i)} - x_t)^2} \tag{3.19}$$

A classical approach for model checking in a Bayesian setting s using different kinds of posterior predictive checking, where one assumes that replicated data generated under the model should look similar to the observed data if the model fits(Gelman et al., 2014). By simulating new data from the joint posterior predictive distribution one can compare different summarizing test quantities to the same quantities in the observed data. This replicated data could be viewed as data that could have been observed, or as data that could be seen in the future, if the estimated model in fact was the one producing the observed data, that is if the model is correctly specified. This replicated data is generated from the posterior predictive distribution, given below.

$$p(y^{rep}|y) = \int p(y^{rep}|\theta)p(\theta|y)d\theta \tag{3.20}$$

where θ contains all parameters in the model, and therefore consists of all states of the latent variable as well as possible house effects or other unkown parameters Replicating the data is in practice done by using each simulation from the posterior distribution and using the estimated parameter values from that simulation to generate a new replicated data set. Using the basic dynamic linear model (3.10) as example, a replicate of the result for party k in each poll y_{ki} is drawn from a normal distribution with the estimated value of the latent state at the time of polling as the mean and the variance for the poll used when modelling. This is repeated for each sample from the posterior distribution. The replicated values are then used to compare test quantities, which capture certain distribution characteristics, with the same measurements in the observed data. Commonly used test quantities are the minimum and maximum value in the observed and replicated data, which can discover how well the posterior predictive distribution catches outliers and tail properties. Another test quantity is the mean, which is used to investigate if the estimates posterior distribution is peaking around the same value as the observed data. The final test quantity used in this thesis will be the variance, which will be used to investigate the spread of the distribution of the observed and the replicated data. These four measurements is a way of investigating if the replicated data shows similar properties in regards to the test quantities to the observed data, and therefore if it is probable that they are generated from the same distribution (Gelman et al., 2014).

A posterior predictive p-value, also called Bayesian p-value or tail-area probability, of the test quantities of the observed data and the replications from the posterior predictive distribution is then calculated. The value of the posterior predictive p-value is the probability that the replicated data is more extreme than the observed data. If the posterior predictive p-value is 1 it means that all test quantities from the replicated data is more extreme than in the observed data, while 0 indicates that none of the test quantities from the replicated data is more extreme. Therefore values close to 1 or 0 indicate a poorly fitting model, since it is desirable to have more extreme values in around half of the replications. The posterior predictive p-value is calculated in formula 3.21 presented below, where $T(y, \theta)$ represented the chosen test quantity (Gelman et al., 2014).

$$p - values_B = Pr(T(y^{rep}, \theta) \geq T(y, \theta)) \tag{3.21}$$

In Table 3.2 the formulas for the test quantities in the posterior predictive checking used in this thesis is presented.

Table 3.2.: Formulas for posterior predictive test quantities.

Test quantity	Formula
Minimum	$Pr(y_{(1)}^{rep} \geq y_{(1)})$
Maximum	$Pr(y_{(N)}^{rep} \geq y_{(N)})$
Mean	$Pr(\frac{1}{N}\sum_{i=1}^{N} y_i^{rep} \geq \frac{1}{N}\sum_{i=1}^{N} y_i)$
Variance	$Pr(\frac{1}{N}\sum_{i=1}^{N}(y_i^{rep} - \bar{y}^{rep})^2 \geq \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2$

3.3. Using periodically collected data in a dynamic linear model

One of the objectives of the thesis is to investigate the effects of handling the issue with periodically collected data. Three different approaches of dealing with this will be tested and evaluated. This is interesting to study since dynamic linear models are often used when one observed data point corresponds with one latent state, which is not necessarily true when dealing with periodically collected political polling data. The techniques reflect different ways of how to incorporate the observed values when modelling. This is not a straightforward procedure since the data collection takes place during an interval time period t. The poll is published after the data collection has taken place. Therefore, it is not necessarily true that the state of the latent variable is the same at the publishing date as it was when the data was collected. It is not even certain that the state of the latent variable is the same during the collection period. This is a challenge for which multiple solutions are plausible, and the effects of different solutions have so far not been investigated in the domain.

The first way to do this is to use the date in the middle of the collection period $p = \{d, d+1, \ldots, D\}$ as the representative time point t^* for when the poll should enter the model. Thus, a poll y_i enters the model as y_{t^*} where

$$t^* = \frac{D - d}{2} \tag{3.22}$$

Letting a time point in the middle of the collection period represent the polling result would still indicate that all the data is collected at one time point, rather than over a time period, but this time point is at least within the actual period when the data was collected which is not the case if the poll would be introduced in the model at the publishing date. This data pre-processing technique will be referred to as data pre-processing technique 1 throughout the rest of the thesis.

Most of the polling houses, whose polls have been used in this thesis, state that their responses are distributed equally over the collection period, meaning that approximately the same number of responses is collected every day. Thus, one could use this knowledge to process the data by dividing the sample size N_p with the number of days in the collection period $p = \{1, \ldots, P\}$, while making the rather strong assumption that the proportion of vote intention for each party is constant during that time period in the poll. This is equivalent of saying that an original poll Y_i is the sum of multiple smaller polls y_i with the same result conducted once a day by the same house during a certain time period.

$$Y_i = \sum_{p=1}^{P} y_p n_p \tag{3.23}$$

where $n_p = N_i/P$. Thus, if a party had 30 percent of the votes in a poll consisting of 1000 respondents, where data was collected during a 5-day period, the poll would be represented by 5 polls where the party had 30 percent but where the sample of each poll was 200. This kind of solution would lead to greater uncertainty of the result of the poll at each given day, since it could be seen to originate from a smaller sample. This data pre-processing technique will be referred to as data pre-processing technique 2 throughout the rest of the thesis. This is similar to the benchmark model, but the average of the latent states are used rather than the outcome of the polls.

To avoid the assumption made in data pre-processing technique 2 regarding the state of vote intention being equal each day of the collection period a third approach is investigated by using the average of the latent states the days of the collection period. This is intuitive since only the sum of the vote intention proportion of a political party for the whole collection period is known. This will be referred to as data pre-processing technique 3 throughout the rest of the thesis, and can be described as in formula 3.24 below. y_{ki} is the outcome of a poll with the collection period $p = \{1, \ldots, P\}$ for party k and x_{kp} is the latent state at each day p for party k in that period. Theoretically this approach can lead to over- or underestimations of the variance, since a sum of binomial distributions is only binomial if the 'probability of success' is constant. Here the proportion is allowed, and even assumed, to vary slightly over the collection period. The effect of issues connected with this should be kept in mind when analysing the result.

$$y_i = \frac{\sum_{p=1}^{P} x_p}{P} \tag{3.24}$$

3.3.1. Simulation studies

Two small simulation studies will be conducted to see how well these three data techniques work in theory to capture the state of a latent variable. In the first simulation study 100 states of a latent variable is drawn from a normal distribution with the variance set to 0.01, where the latent variable has the Markov property. This will represent the 'latent' state for one fictional political party, with the initial vote intention proportion set to 0.5 in an attempt to avoid issues with normal approximation of the binomial distribution. These 'latent' states are used to simulate poll results conducted during different length intervals in a 100-day period. Poll y_i is assumed to have collection period consisting of P days. The numbers of responses in the poll for each day in the collection period is randomly selected, but are restricted to sum to the total sample size 1000. Each poll y_i is generated by sampling from the binomial distribution, where the sample size for each day in the collection period is used as the parameter reflecting 'the number of trials' and the value of the latent variable at each day of the collection period is used as the probability of success. Using a known 'latent' states allows checking of how well the model actually estimates the dynamics of the latent variables in this basic dynamic linear model. The data used in this simulation study is generated by:

$$y_t \sim Binomial(n_t, x_t) \tag{3.25}$$

$$x_t \sim N(x_{t-1}, 0.01)$$

$$x_1 = 0.5$$

$$n_t = 1000$$

When the data is simulated the basic dynamic linear model (3.9) is used to estimate the $x_{1:100} = \{x_1, ..., x_{100}\}$, which can be compared to the known 'latent' series $x_{1:100}$ used to generate the polls $y_{1:100}$. That is, y_t in the model is either represented as y_{t^*}, Y_i or y_p explained 3.22, 3.23 and 3.24.

This simulation study is reproduced generating both the latent variables and the polls from the Dirichlet distribution. The same set-up as described above is used, with the difference that 9 different 'parties' are simulated at once, where the initial values used are the election results in 2014. The value of the concentration parameter is chosen by trial and error when generating the known 'latent' states to keep the probabilities for all 'parties' away from 0, but still allowing for some volatility of the 'latent' states. When modelling the concentration parameter is treated as unknown and estimated by the model just as when modelling with the political polling data. The model used to estimate the latent state is the basic Dirichlet-Dirichlet model in 3.15. The data used is simulated as:

$$\boldsymbol{y}_t \sim Dirichlet(\boldsymbol{x}_t, n_t) \tag{3.26}$$

$$\boldsymbol{x}_t \sim Dirichlet(\boldsymbol{x}_{t-1}, 2000)$$

The results from the simulation studies will be evaluated by calculating how many of the latent states that are covered by the 95% central credible bands, along with the RMSE between the expected value of the posterior distributions of the estimated latent states and the known values of the 'hidden' states.

4. Results

4.1. Simulation studies of techniques for handling periodically collected data

4.1.1. Gaussian distributed latent states

The expected values of the posterior of the 'latent' variable using data pre-processing technique 1, 2 and 3 is presented in Figure 4.1 together with the known values of the 'latent' states. Firstly, the simulated polls are somewhat noisy and do not necessarily capture the true value of the 'latent' time series in a representative way. This is however similar to the assumptions of the dynamic linear model where the data is assumed to be a noisy measurement of the latent time series, making it a believable comparison to the applied problem.

One can see that using data pre-processing technique 1 leads to slightly more jagged estimation compared to when the other techniques are used. All of the data pre-processing techniques are influenced by the polling results, but the estimation of the latent variable seems to rely more heavily on the polls using data pre-processing technique 1. This is most clear before January and March 2015.

The reliance of the data is even more obvious when studying the 95% central credible bands for the expected value of the latent variable. The certainty in the information in a poll makes the credible bands become narrower around poll results while they widens between polls.

The central credible bands using data pre-processing technique 2 indicates that the uncertainty of the latent state is more similar during the studied time period. But the credible bands are too narrow to capture many of the true 'latent' states. Data pre-processing technique 3 is almost a mixture of the other two, where the credible bands narrower when there are several polling results available but also has the widest bands in parts of the time series indicating a greater uncertainty of the estimates when data is scarce. This indicates a high certainty in the polling results.

By using data pre-processing technique 1 92% of the true 'latent' states are covered by the credible bands, while the corresponding percentage for data pre-processing technique 2 is 89% and 96% for data-preprocessing technique 3. One can therefore conclude that most of the true values of the latent states are covered by the credible bands using data pre-processing technique 3. It is not very promising for technique

2 where only 89 out of 100 states are covered by the credible bands since 95 % of the posterior distribution is within in the bands. However, data pre-processing technique 2 would probably work more satisfactorily if the dynamics of the latent state where more smooth. Since the dynamics of the vote intentions are unknown it might be too restrictive to use this data pre-processing technique on political polling data.

Further, the issue involving over- or underestimated variance that might occur with a nonconstant success probability, as mentioned in section 3.3, do not seem to influence the results in an important way. This conclusion is strengthened by trials of estimating the outcome of the poll for each day in the collection period individually instead of as the sum of the days within the period, evading possible problems with binomial sum variance inequality, with very similar results.

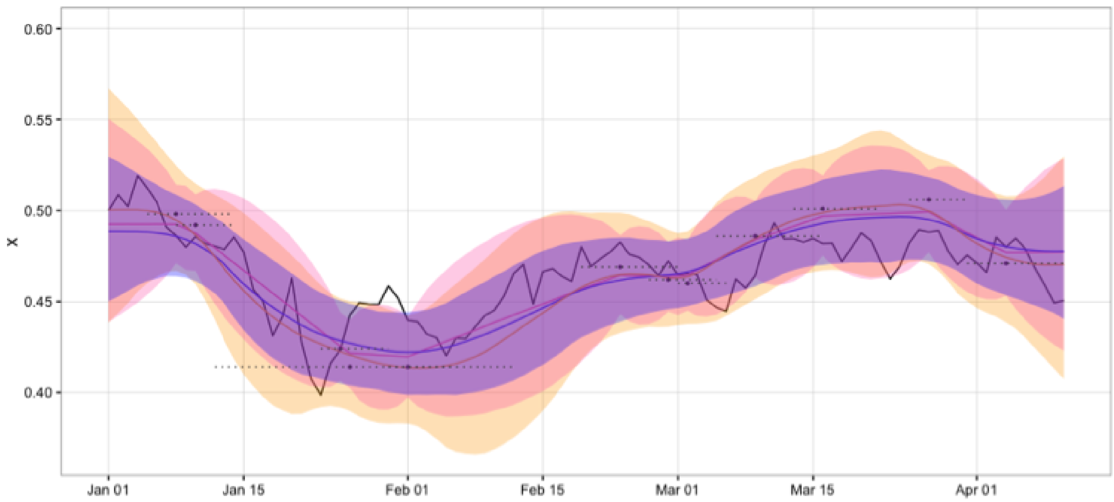

Figure 4.1.: Known values of the 'latent' variable with the expected value of the posterior distribution of 'latent' variable data pre-processing technique 1 (pink), 2 (blue) and 3 (orange) in a basic DLM. Points are the dates polls are entered into the model using data pre-processing technique 1, dotted horizontal lines are the whole collection period, bands are 95% central credible bands.

The RMSE for each of the data pre-processing techniques is presented in Table 4.1 below. Using data pre-processing technique 1 it is 0.0166, while it is slightly lower when using data pre-processing technique 3, 0.0163 and data pre-processing technique 2 gives an RMSE of 0.0182. Thus, the RMSE confirms the results indicating that data pre-processing 3 is the most appropiate to use followed by data pre-processing technique 2, from above.

Table 4.1.: RMSE between expected value of the posterior distribution of the latent states and the true 'latent' states.

Data pre-processing technique	RMSE
1	0.0166
2	0.0182
3	0.0163

Thus, both the visual inspection and the RMSE indicates that data pre-processing technique 3 might be preferable in this setting. The simulation study above were replicated several times using different initial values for the latent state with similar results, where data pre-processing technique 3 always seemed the most promising to use in the overwhelming majority of the cases.

4.1.2. Dirichlet distributed latent states

In Figure 4.2 below one can see the estimations for the first 4 categories out of the 9 used in this small simulation study, the rest is found in appendix B.5 – Figure B.10. The results are similar to as when the latent variable was assumed to be Gaussian, namely that the uncertainty in the estimation of the latent variable using data pre-processing technique 1 and 3 increases between polls and that the uncertainty using data pre-processing technique 2 is equally smooth throughout the time series.

The issues associated with data pre-processing technique 2 is clear if one studies the bottom left plot in Figure 4.2 the first two weeks in February, where many of the latent states are outside the credible bands since they are so narrow.

None of the data pre-processing techniques captures 95% of the latent states within its credible bands for all the 'parties'. However, as could be seen in Table 4.2 below data pre-processing technique 3 has the highest percentage of 'latent' states within the credible intervals for most of the parties. The lowest percentage of 'latent' states within the credible bands is reported for data pre-processing technique 2, where only 64 per cent of the 'latent' states is captured for 'party' 3. This is consistent with the results from the earlier simulation study.

Data pre-processing technique 1 has the equal number of latent states within its intervals as technique 3 for 'party' 5 and 6, and is the technique with most latent states within its credible bands for 'party' 9. This can be seen in Table 4.2.

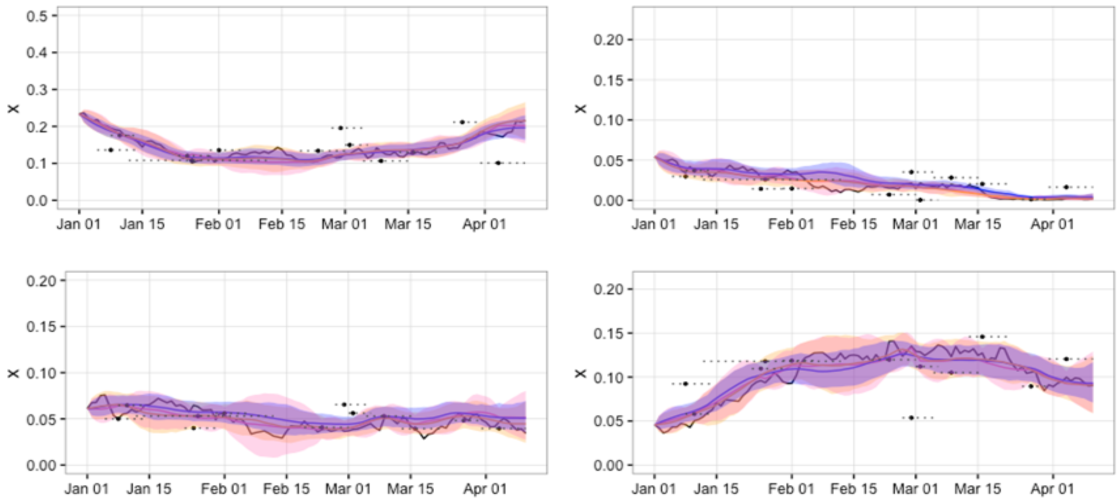

Figure 4.2.: Known values of the 'latent' variable with the expected values of the posterior distributions of the latent state for the first 4 'parties' using data pre-processing technique 1 (pink), 2 (blue) and 3 (orange). Points are plotted at the dates polls are entered into the model using data pre-processing technique 1, dotted horizontal lines show the whole collection period, bands are 95% credible bands.

Further, in Figure 4.2 it can be seen that data pre-processing technique 2 has issues with capturing the 'latent' state when it is more volatile, or in periods with low data frequency.

Table 4.2.: Proportions of the 'latent' states covered by the 95% central credible bands.

	Party								
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Technique 1	0.95	0.94	0.88	0.95	0.97	0.85	0.95	0.90	0.95
Technique 2	0.96	0.77	0.64	0.83	0.79	0.77	0.67	0.87	0.86
Technique 3	0.98	0.97	0.89	0.97	0.97	0.85	0.97	0.95	0.90

The RMSEs when using data pre-processing technique 1 are the smallest for 4 out of the 9 'parties' and also has the lowest average RMSE. Data pre-processing technique 3 also has the lowest RMSE for 4 of the 'parties' while data pre-processing technique 2 has the lowest RMSE for only one of the 'parties'. Data pre-processing technique 2 yields the lowest RMSE when the 'latent' state is very smooth, but is clearly outperformed by the other techniques when the variability increases. These results confirm the previous statements that the assumption made in data pre-processing technique 2 of the result of the poll being equal for each day of the collection period is too strong.

This is similar to the Gaussian case, where the certainty of the estimations is too high to capture fast real changes in the latent state.

Table 4.3.: RMSE between expected value of the posterior distribution of the latent state and the true 'latent' state for data pre-processing technique 1, 2 and 3 for a multi-party system consisting of 9 parties. Italic marks the lowest value.

	RMSE for data pre-processing technique		
'Party' number	**1**	**2**	**3**
1	0.0132	*0.0103*	0.0117
2	0.0080	0.0095	*0.0077*
3	*0.0054*	0.0076	0.0056
4	0.0097	0.0110	*0.0089*
5	*0.0147*	0.0205	0.0174
6	0.0161	0.0159	*0.0158*
7	*0.0064*	0.0088	0.0066
8	0.0105	0.0094	*0.0102*
9	*0.0056*	0.0064	0.0072
Average	*0.0100*	0.0110	0.0101

The lack of variability when using data pre-processing technique 2 seems favourable in this particular case with a smooth latent time series. However, data pre-processing technique 3 covers most of the latent states within the 95% credible bands for most of the 'parties'.

Further, looking in appendix (B.5 – Figure B.11) one can see the posterior distributions the concentration parameters estimations in the models using the different data pre-processing techniques. This value was 2000 when generating the 'latent' state, which is covered by the posterior distributions of the parameters using all three methods. The highest value of this parameter is obtained using data-preprocessing technique 2, which is an additional explanation to the smoothness of the estimations latent states and the corresponding credible bands.

This simulated was replicated several times with a similar result. Since the 'latent' states are best covered by data pre-processing technique 3 along and yields the smallest RMSE for many of the 'parties', this will be the technique used for all the Dirichlet-Dirichlet models presented in this thesis.

4.2. Benchmark model

As a result of the data pre-processing one can see that the estimated value of the latent variable form a rather smooth line along the mean of the polling result from the houses included in the model. By adding the data together and using the combined results for the polls conducted in the same month, the certainty of a specific result is very high. The effects of excluding pollsters that use web-panels become obvious for the Swedish Democrats, which are estimated quite low through out the whole period. The width of the confidence bands heavily increases in mid 2008 and 2009, probably due to low data frequency.

The confidence bands for the data for the Swedish Democrats are narrower than for the other political parties, indicating a greater certainty about the estimation for this party especially in the beginning of the time series. This is not consistent with pollsters having greater issues with capturing the vote intention for this party compared to other but can be a consequence of the use of binomial variance formula, which is maximized when the 'probability of success' is 0.5. A plot for the Swedish Democrats along side the Christian Democrats for comparison of the bandwidth can be seen in Figure 4.3 below.

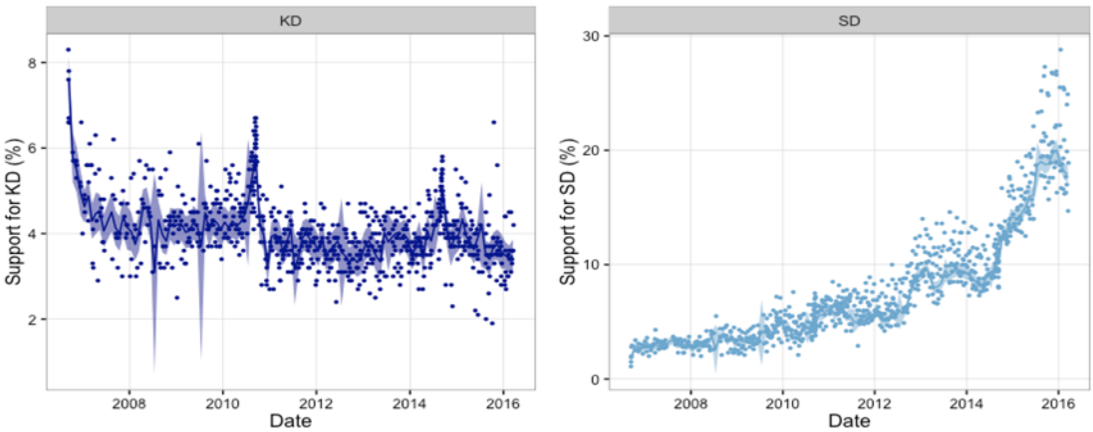

Figure 4.3.: Expected values of the posterior distributions of the latent variable where the lines are the respective benchmark models, polling data as points and ribbon for the 95% credible bands. KD(left) and SD(right).

The next step of evaluating the benchmark model is to see how it performs in regards to estimating the election outcome in 2010 and 2014, which is presented in Table 4.4 below. One would expect that the model copes rather well if not all polls are very wrong. However, studying the table below it can be seen that the model misses the majority of the election results, but that the model performs better in 2010 compared to 2014. Also, the model does better when estimating the election results for the Alliance parties than the left-winged parties and the Swedish Democrats.

Furthermore, the model shows the tendency of overestimating the Green Party and the Left Party while underestimating the Swedish Democrats. The model does not manage to capture the election result for the Social Democrats in any of the elections, which is somewhat surprising since the party have a tradition of becoming the largest political party and do not show greater variability over time than other parties. The only party for which election results are covered by the credible bands in 2014 is the Centre Party.

Studying the PRMSE it can be seen that it is higher for the two major parties, M and S, which once again might be explained by the use of the binomial variance since M and S have vote intention proportions closest to the value maximizing the variance formula. Studying the PRMSE measurement below one can see a curious result, namely that it has a lower value for M and C when the credible bands do not capture the true election value compared to when the true value are within the bands even if the width of the credible bands are the same. This might be caused either by the bias between the simulated values from the posterior and the election results or the variance amongst draws from the posterior.

Table 4.4.: Model comparison between point estimate of benchmark model and election results with 95% credible bands. Bands containing the election results are marked in bold. PRMSE for the posterior distribution and the election result to the right. Election year 2010 and 2014.

Party	Year	Election results	$E(x_{Election})$	95% CB	PRMSE
M	2010	30.1	30.8	**30.1;31.4**	0.0516
	2014	23.3	21.8	21.2;22.5	0.0475
L	2010	7.1	7.4	**7.1;7.8**	0.0080
	2014	5.4	6.8	6.4;7.2	0.0098
KD	2010	5.6	5.9	**5.5;6.2**	0.0225
	2014	4.6	5.2	4.8;5.5	0.0254
C	2010	6.6	6.1	5.7;6.4	0.0071
	2014	6.1	6.3	**5.9;6.6**	0.0109
S	2010	30.7	29.0	28.3;29.6	0.0868
	2014	31.0	29.9	29.2;30.6	0.0647
MP	2010	7.3	8.3	7.9;8.6	0.0052
	2014	6.9	9.4	8.9;9.1	0.0094
V	2010	5.6	6.0	**5.6;6.3**	0.0148
	2014	5.7	6.8	6.4;7.2	0.0196
SD	2010	5.7	5.0	4.7;5.3	0.0240
	2014	12.9	9.8	9.3;10.4	0.0235

The result from the posterior checking is presented in the appendix B.1 - Table B.1 where it can be seen that many of the values are close to either zero or one. The inability to capture extreme values might be an effect of using the aggregated monthly mean in the model, which can also explain why the means in the replicated values are more similar to the observed means than the other test quantitates.

None of the replications contain negative values or values greater than one. This is expected since none of the monthly average for either party is close to being one or negative.

Table 4.5.: Summary model evaluation measurements benchmark model.

	Benchmark model
Number of election result covered by credible bands 2010	4 of 8
Number of election result covered by credible bands 2014	1 of 8
MAD 2010	0.0070
MAD 2014	0.0144
RMSE 2010	0.0083
RMSE 2014	0.0169
Average PRMSE 2010	0.0275
Average PRMSE 2014	0.0264
Number of Bayesian p-values between 0.1 and 0.9	8 of 32
Number of negative values	0 of 32
Number of values > 1	0 of 32

In the benchmark model it is assumed that the true value of vote intention corresponds with the polls coming from pollsters that do not use web panels as data collection techniques. From the results presented above, and summarized in Table 4.5, it can be see that this yields a reasonable working model but with room for improvement especially in regards to comparison between the draws from the posterior on Election Day and the true election result.

The fact that the benchmark also performs very poorly trying to capture the election results for 2014 would indicate that the mean of the polls was dissimilar to the election result, which might be due to bias in the polls or fast changes in the latent variable. Finding a model that can handle this better will be the challenge for the rest of this thesis.

However, if the reason for the benchmark models issues with capturing the latent state on the Election Day 2014 is due to unusual behaviour of vote intention before the election, the rest of the models will likely encounter similar difficulties.

4.3. The basic dynamic linear model

Since the results from the simulation study in section 4.1.1 so clearly indicates that data pre-processing technique 3 is the most appropriate to use, focus will be on these results and they will be the only ones presented in Figure 4.4 below. The plots for the rest of the parties can be found in appendix B.2 - Figure B.4. However, some results from the different data pre-processing techniques will be presented to enable comparisons of the techniques performance on political polling data.

Figure 4.4.: Expected values of the posterior distributions for the latent variable as a line using data process technique 3 in a dynamic linear model, polling data as points and ribbon for the 95% credible band. S(left) and SD(right).

Firstly, it is obvious that the expected values of posteriors of the latent time corresponds rather well with the polling data, indicating that vote intention at least to some extent is captured by the polls.

It can be seen that the model follows the polling results from the pollsters that do not use web panels for the Swedish Democrats in the late 2015 and early 2016, rather than the poll results yielding higher proportion, even if they have been less close to the election outcome for the party. Thus it could be seen as a deficiency of the model that the expected value of the posterior distributions of the latent state is closer to the lower estimated results. However, the estimation of the latent states in the end of the time series appears higher than in the benchmark model, a results from adding houses using web-panels.

Since the polling results during that time period show a high discrepancy between houses it might be reasonable to say that the model estimates the latent states to be somewhere in-between. However, the model show a high certainty that the latent states are somewhere in-between the high and low poll results, indicated by the width of the confidence bands, suggesting that all polls are wrong regarding the vote intention proportions for SD.

The more probable case is perhaps that the true state of vote intention is either captured by the houses using web-panels or by the houses using other data collection techniques. Instead of the model estimating the latent state to be almost somewhere in-between the poll results it would be more desirable for the model interpreting the divergance of polling results as increased uncertainty of the state of the latent variable, leading to wider credible bands.

The model captures the more extreme changes in political opinion very well, for example the previously mentioned Juholt scandal for the Social Democrats in the end of 2011.

Another general result seen in the plots is that the bands are more narrow when fast changes in vote intention seem to occur, while they become wider in periods when the party sympathies are less volatile. This might be explained by the decreasing frequency in polling between elections compared to during election campaigning or this can simply be a visualization problem. When plotting only the period for the rapid change it can be seen that the credible bands occur less narrow than in the Figure 4.4 above, which would support the explanation of it being a visualization problem.

A comparison of the models ability to capture true election results is presented in the Table 4.6 below. The table includes a comparison of the general election 2010 and 2014 and the expected values of the posterior distributions of the latent variable for each party. All of the election results except for MP are within the 95 % credible bands for 2010 using data pre-processing technique 3, while the result is less accurate in 2014 where only three of the bands covers the election results. This is an improvement from the benchmark model, but there is still room for further improvements.

Using data pre-process technique 1 yields equally accurate result in 2010, with seven of the election results included within the credible bands. Thus, using data pre-processing technique 1 and 3 seems superior to the benchmark regarding the posterior covering the true election result for both years. Further, several of the election results are just outside the 95 percent credible bands for the model using data pre-processing technique 1 and 3. Using data pre-processing technique 2 yields less accurate results.

The PRMSEs are quite similar for all three data pre-processing techniques. Neither of the models have an overall obviously lower PRMSE compare to the benchmark model. It can be seen that the posterior distribution at Election Day of the Moderates and the Social Democrats seems less similar to the election results than the other parties, which is as previously mentioned most likely a consequnce of the constant prior used in the model

Just like in the benchmark model it can be seen that the PRMSE in 2014 is lower for the Moderates than in 2010 even though the credible bands is not contain the true value. However, the credible bands are more narrow in 2014, indicating a lower variance of the draws from the posterior distribution which might explain the lower PRMSE. Further, the credible bands just barely miss the true election results in 2014 for the Moderates. The fact that the different evaluation measures of the posterior distributions of the latent states on Election Day show dissimilar results shows the necessity of multiple different ways of assessing the same model, displaying its strengths and weaknesses.

By summing the point estimates for the different political parties for each year one can see that they sum to values lower than 100 for all three data pre-processing techniques for both years. This is not a constraint that is incorporated into the model but it is however a constraint of the nature of the problem at hand.

Table 4.6.: Point estimates and 95% confidence band using data pre-processing technique 3 in basic dynamic linear model. Bands containing the election results are marked in bold. Election year 2010 and 2014.

Party	Year	Election results	$\mathbf{E(x_{Election})}$	95% CB	PRMSE
M	2010	30.1	29.1	**27.2;31.0**	0.0522
	2014	23.3	21.9	20.7;23.2	0.0475
L	2010	7.1	7.3	**6.6;8.1**	0.0061
	2014	5.4	6.4	6.0;6.9	0.0093
KD	2010	5.6	5.8	**5.2;6.4**	0.0126
	2014	4.6	5.0	**4.6;5.4**	0.0109
C	2010	6.6	6.5	**5.8;7.3**	0.0098
	2014	6.1	5.8	**5.3;6.3**	0.0246
S	2010	30.7	29.4	**26.7;32.1**	0.0835
	2014	31.0	30.3	**28.5;32.1**	0.0624
MP	2010	7.3	8.7	7.8;9.6	0.0153
	2014	6.9	9.0	8.2;9.8	0.0173
V	2010	5.6	6.1	**5.4;6.7**	0.0131
	2014	5.7	6.8	6.1;7.4	0.0213
SD	2010	5.7	5.1	**4.2;5.9**	0.0233
	2014	12.9	10.1	9.5;10.8	0.0251

When looking at the evaluation measurements for the model presented in appendix B.2 - Table B.4 it can be seen that the posterior predictive distribution do not seem to share the characteristics of the observed data, since almost all Bayesian p-values presented in the appendix are close to zero or one. This is an obvious sign of that the observed data was not generated by the posterior distribution suggested by the model. Therefore it is motivated to add components to the model that might increase the variability of the model such as house effects.

The main exception from the Bayesian p-values being zero or one is when using data pre-processing technique 1 for the Moderates where the minimum, maximum and the mean value in the generated data are closer to observed data. Data pre-processing techniques 2 overall generates the fewest Bayesian p-values further away from 0 and 1. That indicates that the model is more likely to generate the observed data using data pre-processing technique 1 or 3. These results also show the advantage of the benchmark model, which seem to capture one feature of the posterior very well.

Further, the basic DLM using data pre-processing technique 3 have lower MAD and RMSE for both years. All three data pre-processing methods yields a lower RMSE for 2014 than the benchmark model, while data pre-processing technique 1 and 2 yields higher values in 2010. The MAD values are either higher than or equal to the benchmark for both years.

Table 4.7 below show a summary of the different evaluation measurements using the three techniques as well as the benchmark model for comparison

Table 4.7.: Summary model evaluation measurements basic dynamic linear model with three different data pre-processing techniques.

	Data pre-processing technique: 1	2	3	Benchmark model
Number of election result covered by credible bands 2010	7 of 8	1 of 8	7 of 8	4 of 8
Number of election result covered by credible bands 2014	2 of 8	1 of 8	3 of 8	1 of 8
MAD 2010	0.0076	0.0099	0.0066	0.0070
MAD 2014	0.0144	0.0115	0.0123	0.0144
RMSE 2010	0.0094	0.0111	0.0082	0.0083
RMSE 2014	0.0159	0.0139	0.0146	0.0169
Average PRMSE 2010	0.0272	0.0269	0.0270	0.0275
Average PRMSE 2014	0.0266	0.0325	0.0273	0.0264
Number of Bayesian p-values between 0.1 and 0.9	5 of 32	0 of 32	2 of 32	8 of 32
Number of negative values	0 of 32	0 of 32	0 of 32	0 of 32
Number of values > 1	0 of 32	0 of 32	0 of 32	0 of 32

Using data pre-processing technique 2 and 3 the expected values of the posterior distributions of the different states of the latent variable have an appearance that is more similar to what would be expected and therefore seem as the more attractive way of pre-process the data. The volatility shown in the time series plots using data pre-processing technique 1 is simply very unlikely keeping in mind the knowledge about Swedish voters, which do not suggest the strong oscillating behaviour obtained by the model using this data pre-processing technique.

It can be concluded that the model evaluation measurements indicates that better results are obtained by using data pre-processing technique 1 or 3, while the more realistic results in term of behaviour of the latent variable are achieved by using data pre-processing technique 2 and 3. These combined results indicate that data pre-processing technique 3 is preferred in the applied case as well as the theoretical and will be used in the rest of the univariate Gaussian models.

4.4. Dynamic linear model with time invariant house effects

The results from using the three models including time invariant house effects in different ways are presented in the following section, starting with a visual inspection.

Using additive house effects on the mean yields similar results to when using a basic DLM, but with a slight increased smoothness of the expected values of the latent variable, seen in Figure 4.5 below. This increased smoothness is the most noticeable when studying the peaks in the expected value of the posterior distributions of the latent variable, where the peaks are not as high as in the basic model. This is an indication that some of the more extreme changes in the observed values can be attributed to the fact that polls are conducted by different houses rather than actual changes in the latent variable.

It can also be seen that the expected values have shifted either slightly upwards or downwards overall for most of the parties. This would be even more apparent without the restraint of the house effects summing to zero. Removing the constraint creates a poorly working model with issues of identification inbetween elections. This makes sense since it is the inclusion of election results that identifies the model, but only around the time of the election. If the house effects for a party are positive it means that the expected value of the latent variable will have been shifted downwards compared to the basic model.

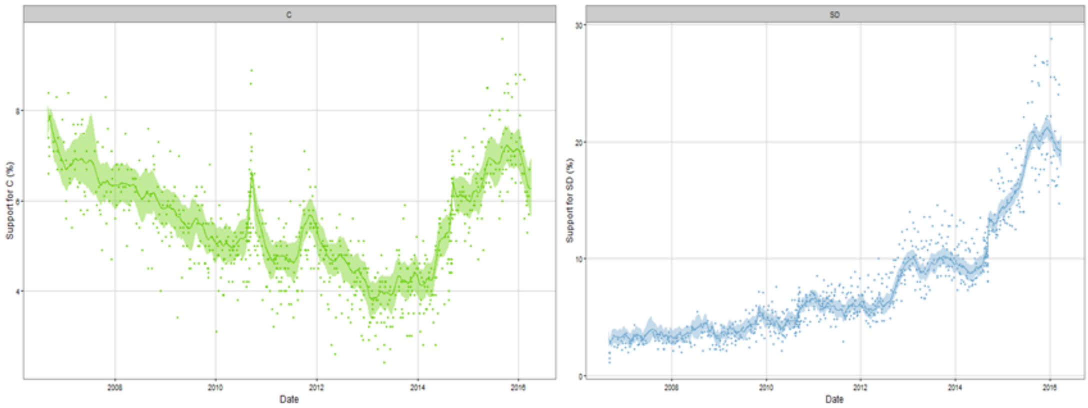

Figure 4.5.: Expected values of the posterior distributions for the latent variable as a line in additive house effects on mean model, polling data as points and ribbon for 95% credible band. C(left) and SD(right).

The width of the credible bands are increased for all political parties when multiplicative house effects in the variance are used, indicating that more variability of the latent variable has been introduced. The increased uncertainty of the latent variable is especially apparent in intervals with a low data frequency early in the time series. The lack of data was noticeable in the basic dynamic linear model as well, visible in Figure 4.6, but studying the plots for the Centre Party and the Swedish Democrats one can see that the upper limit of the 95% credible band peaks twice in 2007 and 2008 in a more apparent way than when using additive house effects on the mean.

It can be seen that the model just like in the basic DLM estimate the latent state of the Swedish Democrats to be almost somewhere in-between the polling results of web-panels and the polls using other data collection techniques and the width of the credible bands do not increase noticeable. Thus, the use of house effects does not lead to a model that capture an increased variability between polls.

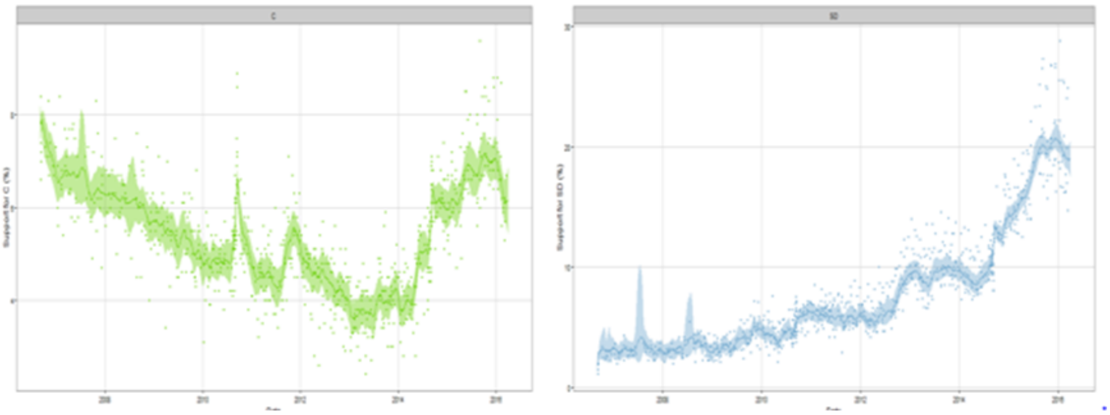

Figure 4.6.: Expected values of the posterior distributions for the latent variable as a line in multiplicative house effects on variance model, polling data as points and ribbon for 95% credible band. C(left) and SD(right).

The results when modelling the house effects as additive on the variance are visually similar to the ones obtained when using an additive house effect on the mean. Studying the plot for Centre Party it can be seen that the credible bands are more equally wide though out the time series than in the two other approaches, as seen in Figure 4.7. Looking closely at the plots for the Swedish Democrats it can be seen that that the credible bands are a bit wider than when the house effects were added to the mean, indicating the increased variability of the model. However, this increased variability is expected since the prior used in the model only can generate positive values.

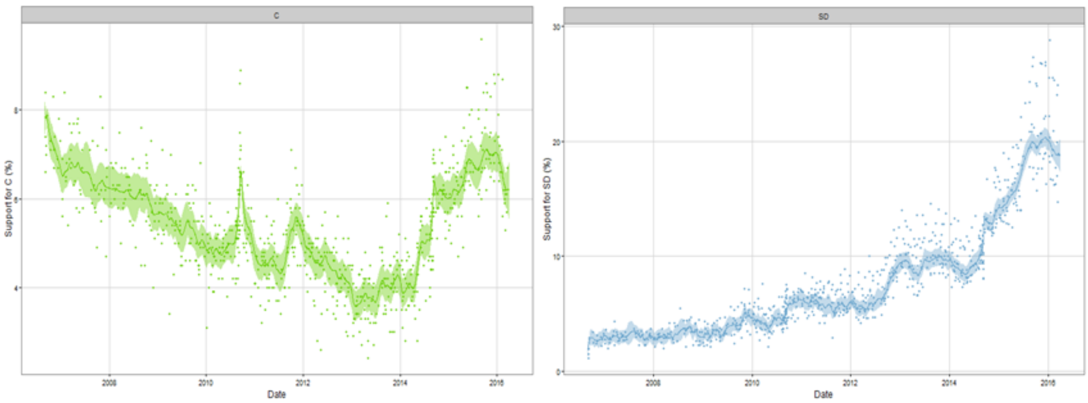

Figure 4.7.: Expected values of the posterior distributions for the latent variable as a line in additive house effects on variance model, polling data as points with ribbon for 95% credible band. C(left) and SD(right).

Modelling the house effects as multiplicative on the mean generated very poor results, and is therefore not presented in this thesis. Experiments regarding distributions and restrictions were made trying to obtain a better working model, such as trying different prior distributions of the parameter and restrictions on sums. However, all of these experiments were unsuccessful since the connection between the measurement equation and the dynamic equation seems to be lost using multiplicative house effects on the mean. That is, there were issues with identifying the model yielding posteriors too heavily influenced by the priors compared to the information in the data. As previously mentioned this becomes an issue in-between elections since the inclusion of election results only identifies the model close the elections.

The posterior distribution for the election days becomes skewed when modelling the house effects as additive on the mean, yielding expected values of the latent variable far over or under the true election result. This is especially obvious for the Moderates, which is grossly underestimated and the credible band is very wide. However, the model covers equally many true election results as the benchmark model in both 2010 and 2014.

The number of credible bands covering the true election result is more balanced between 2010 and 2014 when using multiplicative house effects on variance compared to the basic DLM, where six of the results are covered in 2010 and four in 2014. Therefore it can also be concluded that the model is better at capturing the election result than the benchmark model. The Green Party is still overestimated in both elections, indicating that the model does not correct the positive bias in the polls for the party in a satisfactory way.

Tables 4.8 and 4.9 below confirms the increased width of the credible bands noted in the visual inspection, since they have increased slightly in around the point estimates of Election Day. Adding the house effects on the variance yields the most accurate results in 2010, and is on par with the model using multiplicative house effects on the variance in 2014. Thus, it is in fact the model that captures most of the election results within the credible bands.

Using multiplicative house effects on the variance yields higher average PRMSE than the previous models, including the benchmark model for 2014, yet again confirming that the PRMSE for SD heavily increase from 2010 to 2014 for all models. This yet again confirms the speculation that the uncertainty in estimations of larger parties is a result from the binomial variance formula used as the constant prior in the models, since SD more than doubled in size between the election 2010 and 2014.

Using additive house effects on both mean and variance yields lower average PRMSE compared to the benchmark model in 2010, but the measurement gets higher values than the benchmark model for the election 2014. The models using house effects on the variance have higher average PRMSEs in 2014 than in 2010. This could be explained by the increased deviation the the election result, reflected in the decrease of number of election results within the 95% credible bands.

It can be seen once again that the PRMSEs are lower for some parties when in true election results is not within the credible bands compared to when the results are captured. This is visible for SD when modelling the house effects as additive on the mean and for V modelling house effects as multiplicative on the variance.

Table 4.8.: Model comparison between point estimate and 95% confidence using different time invariant house effects. Bands containing the election results are marked in bold. Election year 2010.

		Additive on mean			**Multiplicative on variance**			**Additive on variance**		
Party	Election result	$E(x_t)$	95% CB	PRMSE	$E(x_t)$	95% CB	PRMSE	$E(x_t)$	95% CB	PRMSE
M	30.1	22.3	13.5;27.9	0.0651	29.3	**27.3;31.3**	0.0523	29.1	**27.1;31.1**	0.0640
L	7.1	8.0	**7.0;9.3**	0.0033	7.3	**6.5;8.1**	0.0063	7.4	**6.6;8.1**	0.0021
KD	5.6	7.8	6.7;9.8	0.0059	5.8	**5.2;6.5**	0.0128	5.8	**5.2;6.4**	0.0054
C	6.6	8.1	7.2;9.4	0.0083	6.4	**5.8;7.2**	0.0096	6.5	**5.9;7.2**	0.0089
S	30.7	29.3	**26.3;31.8**	0.0329	29.1	**26.7;35.0**	0.0847	29.3	**26.7;32.0**	0.0330
MP	7.3	8.7	**7.5;10.1**	0.0076	8.6	7.7;9.6	0.0151	8.7	7.9;9.7	0.0075
V	5.6	6.7	5.7;8.9	0.0016	6.0	**5.4;6.7**	0.0128	6.0	**5.5;6.6**	0.0022
SD	5.7	7.8	6.11,6.0	0.0062	4.8	3.9;5.5	0.0232	5.1	**4.2;5.9**	0.0062

Table 4.9.: Model comparison between point estimate and 95% confidence band using different time invariant house effects. Bands containing the election results are marked in bold. Election year 2014.

		Additive on mean			**Multiplicative on variance**			**Additive on variance**		
Party	Election result	$E(x_t)$	95% CB	PRMSE	$E(x_t)$	95% CB	PRMSE	$E(x_t)$	95% CB	PRMSE
M	23.3	14.0	6.3;18.7	0.0621	22.0	**20.7;23.3**	0.0482	22.0	**20.6;23.3**	0.0480
L	5.4	7.5	6.7;8.4	0.0190	6.5	6.0;6.9	0.0095	6.5	6;6.9.0	0.0100
KD	4.6	7.1	5.9;9.4	0.0148	5.1	**4.6;5.5**	0.0084	5.0	**4.6;5.6**	0.0100
C	6.1	8.2	7.2;9.6	0.0145	5.9	**5.4;6.5**	0.0152	5.9	**5.5;6.5**	0.0141
S	31.0	29.2	**27.1;31.2**	0.0498	30.2	**28.5;31.9**	0.0558	30.3	**28.3;32.3**	0.0557
MP	6.9	9.2	8.1;10.4	0.0216	8.9	8.1;9.6	0.0195	8.8	7.9;9.7	0.0200
V	5.7	8.5	7.4;10.1	0.0192	6.7	5.9;7.4	0.0083	6.6	5.8;7.4	0.0100
SD	12.9	12	**10.7;14.0**	0.0521	9.9	9.1;10.7	0.0808	9.1	8.7;10.9	0.0819

Only one of the p-values from the posterior predictive check is between 0.1 and 0.9 when using house affects that are additive on the mean, and it is obtained comparing the variance in original and replicated data for the Left Party which is almost 0.5. This means that around half of the replicated data has a is higher variance than in the original data, while the other has a lower variance. Using multiplicative and additive house effects on the variance yields only two Bayesian p-values between 0.1 and 0.9, which is on par to the basic DLM but lower than the benchmark model. Thus, the results from the different ways of incorporating the house effects through the variance of the measurement equation continue to be similar. This is presented in the appendix B3 – Table B.5.

In the summary below one can also see how the previously mentioned width of the credible bands of the latent state at the day of the election in 2010 for the Moderates affects both MAD and RMSE when using additve house effect on the mean. A similar result is obtained in 2014, where the uncertainty of M heavily affects the total average for MAD and RMSE. This makes both measurements much larger than for the other models even if the distance between the election results and the expected values of the posterior distributions are more similar to the results from the other models for the rest of the parties.

Table 4.10.: Summary model evaluation measurements dynamic linear model with three different time invariant house effects.

	Additive on mean	Multiplicative on variance	Additive on variance	Benchmark model
Number of election result covered by credible bands 2010	3 of 8	6 of 8	7 of 8	4 of 8
Number of election result covered by credible bands 2014	2 of 8	4 of 8	4 of 8	1 of 8
MAD 2010	0.0230	0.0070	0.0068	0.0070
MAD 2014	0.0298	0.0124	0.0129	0.0144
RMSE 2010	0.0313	0.0086	0.0084	0.0083
RMSE 2014	0.0385	0.0149	0.0168	0.0169
Average PRMSE 2010	0.0272	0.0269	0.0270	0.0275
Average PRMSE 2014	0.0266	0.0325	0.0273	0.0264
Number of Bayesian p-values between 0.1 and 0.9	1 of 32	2 of 32	2 of 32	8 of 32
Number of negative values	0 of 32	0 of 32	0 of 32	0 of 32
Number of values > 1	0 of 32	0 of 32	0 of 32	0 of 32

The house effects obtained using the different time invariant house effect models is presented in the appendix B.3 – Table B.6. The expected value of the house effects for the Swedish Democrats are positive for all houses using self-recruiting web-panels: Inizio, Sentio, United Minds and YouGov. This simply reflects the fact that the polls coming from houses using self-recruiting web-panels have a higher mean than polls from pollsters using different data collection method. The strongest positive house effect is found for YouGov followed by Sentio.

Furthermore, Novus has in previous works been found to be the most similar to the mean estimation of the parties in general, and the results here show that the point estimate for Novus is very close to zero, or even zero, for all parties. This indicates that the bias in polls from Novus is small. However, the over all sizes of the house effects are very small due to the restriction imposed when modelling indicating that the houses on average have no bias.

The model using multiplicative house effect on the variance shows a similar pattern, namely that the house effects for Sentio and YouGov are high for the Swedish Democrats. If the house effect is higher than 1 it indicates that the sampling design is less effcient than random sampling. Sentio, YouGov, United Minds and SVT all have a large positive house effects for SD, thus indicating that the sampling for these houses are less efficient than what is expected by random sampling. A high value in the multiplicative house effect means that the precision of the poll is decreased and therefore that the variance increase. This indicates that there is some bias in the poll that affect the outcome in the measurement equation rather than real changes in the latent variable. This is not very surprising since self-recruiting web-panels cannot be considered random samples.

The fact that SVT have a less efficient sampling method than the others for all political parties might be explained by the fact that they use cluster sampling which can at best perform on par with random sampling. However, the size of the effect is still somewhat surprising since they only ask people partaking in the election, which would decrease the deviation between vote intention and actual voting, which might have increased the efficiency of the sampling.

Also, the size of the house effects from SVT Valu is interesting since the binomial variance for these exit polls will be low due to the large sample size, which the house effects might try to adjust for. The result for SVT Valu is also interesting since these polls take place in the voting locations at the day of the election and are the only polls in the data when people face the interviewers directly. The direct contact with an interviewer can increase the bias if the question could be considered sensitive.

Novus is estimated to have house effects below 1 indicating a more efficient sampling than expected in random sampling for almost all parties.

Adding the house affects additively to the variance yields very similar estimates of the house effects as when adding them multiplicatively, where the same polling houses increase the variance the most for the same parties. However, the interpretations are less straight forward since they cannot any longer be interpreted as differences in efficiency compared to random sampling. The smaller the value of the house effect, the more the variance is increased compared to when only using the binomal variance used as a constant prior in the models.

The results above indicate that the use of either multiplicative or additive house effects to the variance is the most appropriate approach in the time invariant case.

4.5. Dynamic linear model with time variant house effects

The time variant house effect will be tested on the multiplicative house effects on the variance since it is found as one of the two most appropriate approaches in section 4.4 as well as having a natural interpretation as design effects. The additive house effects on the variance also have the issue with only being able to increase the variance, whereas the multiplicative house effect can decrease the variance as well.

Figure 4.8 show the two parties, KD and SD, where the difference between the results when using time variant multiplicative house effects on the variance and corresponding time invariant effects are the clearest. Using time variant house effects increases the variability of the model, especially when data frequency is low.

This is very visible in 2007 when there is an data gap for almost two months. This is also the first model in which the 95% central credible bands cover negative values for parties with lower proportions of people with the intention to vote for them. The extreme behaviour of the house effects is an indication that the vague prior used allows for too extreme values when data is scarce.

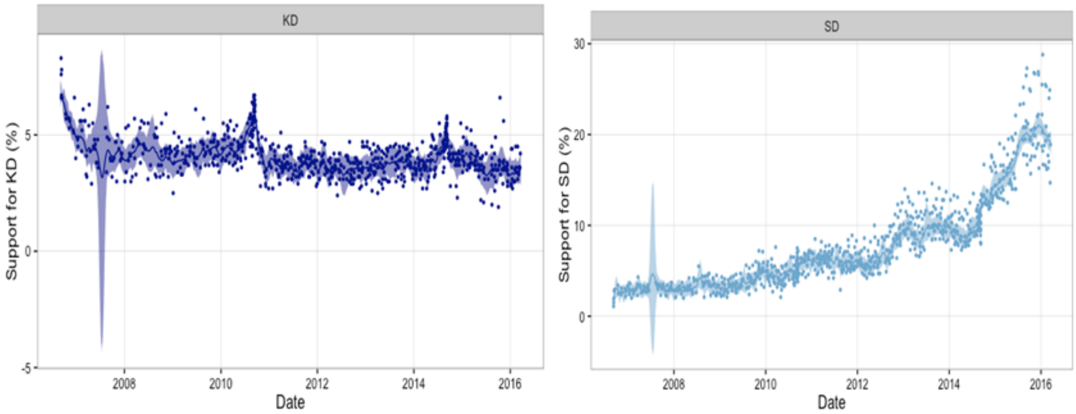

Figure 4.8.: Expected values of the posterior distributions for the latent variable as a line in DLM with time variant multiplicative house effects on the variance, polling data as points and ribbon for 95% credible band. KD(left) and SD(right).

Studying table 4.11 it is clear that the model using time variant house effects is less successful than its time invariant counterpart in regards of capturing the election results for the different parties at Election Day. Half of the credible bands for the parties cover the true election results in 2010, where all of these parties belonged to the right wing Alliance.

In 2014 the results are even poorer where the credible bands cover the election results for only two of the parties. The parties with the highest PRMSEs are as before the two largest parties and MP and SD.

Table 4.11.: Point estimates and 95% confidence band using data pre-processing technique 3 in a DLM with time variant house effects model. Bands containing the election results are marked in bold. Election year 2010 and 2014.

Party	Year	Election results	$E(x_{\mathbf{Election}})$	95% CB	PRMSE
M	2010	30.1	29.0	**28.0;30.1**	0.0122
	2014	23.3	21.8	20.9;22.8	0.0256
L	2010	7.1	7.3	**6.8;7.8**	0.0037
	2014	5.4	6.4	5.7;6.9	0.0081
KD	2010	5.6	5.8	**5.4;6.2**	0.0030
	2014	4.6	5.1	**4.6;5.6**	0.0047
C	2010	6.6	6.4	**5.9;6.9**	0.0091
	2014	6.1	6.0	**5.2;6.9**	0.0066
S	2010	30.7	29.1	27.7;30.5	0.0175
	2014	31.0	30.3	**29.1;31.5**	0.0215
MP	2010	7.3	8.8	8.2;9.3	0.0147
	2014	6.9	8.9	8.0;9.6	0.0177
V	2010	5.6	6.0	5.7;6.4	0.0049
	2014	5.7	6.9	6.2;7.4	0.0096
SD	2010	5.7	4.7	4.2;5.3	0.0102
	2014	12.9	9.9	9.0;10.8	0.0352

As could be seen in Figure 4.8, this is the first model where negative values where generated from the posterior distributions of the latent states. This is captured when new data is replicated from the posterior predictive distribution as well, where negative values were amongst the replicated values for the first time in this thesis. The fact that negative values is possible is a consequence when assuming that the vote intention for each party follow a univariate Gaussian distribution.

However, the amount of negative values amongst the replicated values where so low that it appears to be zero in appendix B.4 – Table B.7, and therefore also in Table 4.12 below. Negative values where only replicated for the Swedish Democrats, where the negative values of vote intention is obtained when replicating polls in the begining of the time series when SD still was a rather small political party having poll results close to zero.

Studying the summarizing Table 4.12 below one can see that fewer Bayesian p-values are between 0.1 and 0.9 than in the benchmark model, but compared to its time invariant counterpart more values from the posterior predictive checks are not zero or one. This indicates that the Gaussian dynamic linear model assuming that house effects evolve as a random walk process are more similar to the model generating the data. Both MAD and RMSE is higher than for the benchmark model 2010 but lower 2014.

Table 4.12.: Summary of model evaluation measurements for DLM with time variant house effects.

	Multiplicative time variant house effect on variance	**Benchmark model**
Number of election result covered by credible bands 2010	4 of 8	4 of 8
Number of election result covered by credible bands 2014	2 of 8	1 of 8
MAD 2010	0.0078	0.0070
MAD 2014	0.0125	0.0144
RMSE 2010	0.0096	0.0083
RMSE 2014	0.0152	0.0169
Average PRMSE 2010	0.0092	0.0275
Average PRMSE 2014	0.0161	0.0264
Number of Bayesian p-values between 0.1 and 0.9	3 of 32	8 of 32
Number of negative values	*0* of 32	0 of 32
Number of values > 1	0 of 32	0 of 32

In appendix B.4 – Figure B.9 it can be seen how the house effects for the different houses and parties changes over time, and from this it can be concluded that they indeed are not constant over the time series. Some of the features mentioned in section 2.2 are captured by the house effects, for example that the house effects of Demoskop for the Moderates are higher in the earlier polls compared to later and Sentio also has high house effects for the party, especially in the beginning and in the end of the time series. The variance of the house effects for the two major parties S and M is more constant over time than for the smaller parties. The interesting pattern that the variability amongst the house effects clearly increase is especially obvious for SD, L and MP, where SD and MP have been two of the parties the models have had the most issues with capturing. An explaination to the increased house effects might be problems with increased non response rates most polling houses are suffering from.

In the appendix B.4 – Table B.8 the expected values of the house effects for each house and parties are presented, where it can be concluded that average house effects are the highest for SD. Compared to the corresponding time invariant estimations of the house effects almost all expected values of the house effects are clearly higher. This along with the other issues with the model might be due to the fact that house effects are estimated on a daily basis through out the time series, even if houses are not producing polls on that day. The vaguness of the prior might explain the extreme behaviour when data is scarce, combined with the extreme behaviour of the house effects in certain time periods could explain the strange expected values of the house effects over all.

To avoid this problem one can either model the house effects on either a monthly or quaterly basis instead of daily, in an attempt of dealing with lack of data and to reduce the number of parameters to estimate in the model. Another approach can be to only estimate the house effects when a poll is conducted. Also, using a $Gamma(1, 1)$ prior of the variance indicates slowly changing house effects, this might have been an erroneous assumption contributing to the flaws of the models. Perhaps a prior allowing for large changes in house effects when methodology changes have occured and otherwise slowly changing would be more appropiate.

Using time variant house effects do not improve the model in an obvious way, but instead yields point estimates at election day that are less similar to election results than the time invariant counterpart. However, the posterior predictive checks indicate that the model is slightly more similar to the one generating the observed data.

4.6. Basic Dirichlet-Dirichlet model

Studying Figure 4.9 of the expected values of the posterior distributions of the latent states and their credible bands it is clear that they are wigglier than in the univariate Gaussian models. The less smooth estimations are reminiscent of the behaviour found when using data pre-processing technique 1 in the Gaussian models, which have been explained previously as being rather unlikely. However, the non-smoothness of the expected values is not as bad, especially for the Swedish Democrats. Also, Figure 4.9 below are also somewhat reminiscant of the results in the benchmark model.

The movement of the latent state is connected with the size of the concentration parameter, and the expected value of posterior distribution of the concentration parameter is estimated to 24325, presented in the appendix B6 - Figure B.13. The higher the value of the concentration parameter the lower variance in the estimations of the latent states, and therefore one can speculate that the parameter is estimated too low.

One lesson learned from the simulation studies was that the size of the concentration parameter heavily affects the stability of the Dirichlet distribution. This can be a consequence of the limitation that the concentration parameter is time invariant. The prior for the concentration parameter cover values much higher than the ones in the posterior distribution, and should not be resticting the model.

The volatility of the estimated values of the latent states is more obvious for the smaller parties than the larger. The expected values of the latent states for the major parties are very similar to the ones obtained in the univariate models, with a smoother latent time series than the ones seen below for the Left Party and the Swedish Democrats.

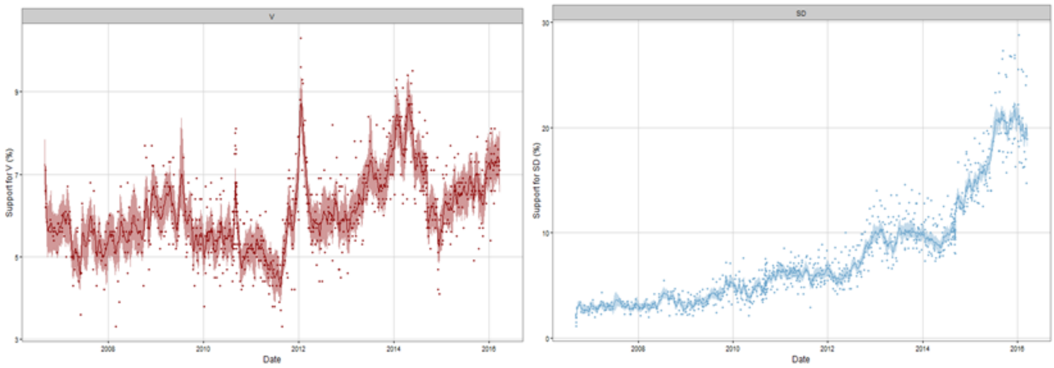

Figure 4.9.: Expected values of the posterior distributions for the latent variable as a line in basic Dirichlet-Dirichlet model, polling data as points and ribbon for 95% credible band. V(left) and SD(right).

Studying Table 4.13 below it is clear that the model once again has issues with capturing the election results of MP within the 95% central credible bands for both elections. It is also the party with the highest PRMSE in 2010, followed by the two largest parties S and M. MP has the second highest PMRSE for the 2014 election after SD. The model also misses the election result for SD in 2010 with 0.2 per cent, and with an even higher margin in 2014. However, the credible bands for the rest of the political parties, as well as the new 'Other parties' category, cover the election results.

The 95% credible bands cover a lot fewer of the election results in 2014 compared to 2010, three out of seven, just like all the previous models. One explanation to this is that the credible bands are narrower for all parties except SD. Also, this is the first model where the PRMSE measurement always is lower when the true election result is within the credible intervals compared to when it is not.

However, the model captures more of the results within the credible bands of the parties then the benchmark model, even if it performs worse than the univariate Gaussian models with house effects on the variance regarding the comparison of the point estimate of the latent state on Election Day and the true election results. Moreover, the average PRMSEs are clearly lower for both 2010 and 2014 compared to the benchmark as well as the univariate Gaussian models with all different house effects.

Many of the election results are very close to the limits of the credible bands, but this basic Dirichlet-Dirichlet model yields a lower MAD and RMSE than the benchmark model in 2014, while the measurements are on par with the benchmark in 2010. This makes sense since more of the election results are within the 95% credible bands.

Table 4.13.: Point estimates and 95% credible band using data pre-processing technique 3 in a basic Dirichlet-Dirichlet model. Bands containing the election results are marked in bold. Election year 2010 and 2014.

Party	Year	Election results	$E(x_{\text{Election}})$	95% CB	PRMSE
M	2010	30.1	28.9	**27.6;30.1**	0.0143
	2014	23.3	21.9	21.1;22.6	0.0155
L	2010	7.1	7.2	**6.5;7.9**	0.0046
	2014	5.4	6.6	**6.1;7.0**	0.0116
KD	2010	5.6	5.9	**5.3;6.6**	0.0050
	2014	4.6	5.2	4.8;5.6	0.0066
C	2010	6.6	7.1	**6.4;7.8**	0.0070
	2014	6.1	6.3	**5.8;6.8**	0.0033
S	2010	30.7	29.6	**28.3;30.9**	0.0132
	2014	31.0	30.3	**29.5;31.2**	0.0085
MP	2010	7.3	9.1	8.3;9.9	0.0180
	2014	6.9	8.9	8.3;9.5	0.0203
V	2010	5.6	6.0	**5.4;6.7**	0.0057
	2014	5.7	6.7	6.2;7.2	0.0101
SD	2010	5.7	4.9	4.3;5.5	0.0088
	2014	12.9	9.9	9.4;10.5	0.0294
O	2010	1.4	1.4	**1.0;1.7**	0.0021
	2014	4.1	4.3	**3.9;4.7**	0.0030

The same number of values from the posterior predictive checks is between 0.1 and 0.9 as for the benchmark model, making the proportion lower for the basic Dirichlet-Dirichlet model since one more party is added. It is mainly the Bayesian p-values for M, L and S that are between 0.1 and 0.9. Also, almost all of the Bayesian p-values for the minimum value is none zero. Thus, the left tail of the distribution is more similar to the observed data.

However, looking in the appendix B.6 – Table B.9 one can see that the number of entries that are not exactly zero or one have decreased compared to the benchmark and the univariate Gaussian models. The number of negative values and the number of estimated proportions above one is of course zero since this is impossible when generating from the Dirichlet distribution where the elements are always positive reals that sum to one.

From the summary presented in Table 4.14 and the other presented results it seems that the model is on par with the univariate models presented in this thesis, but with the obvious benefit of the proportion of vote intention and the outcome of the polls are restricted to sum to 1 and negative values are impossible.

Table 4.14.: Summary model evaluation measurements basic Dirichlet-Dirichlet model

	Basic Dirichlet-Dirichlet model	**Benchmark model**
Number of election result covered by credible bands 2010	7 of 9	4 of 8
Number of election result covered by credible bands 2014	3 of 9	1 of 8
MAD 2010	0.0070	0.0070
MAD 2014	0.0133	0.0144
RMSE 2010	0.0087	0.0083
RMSE 2014	0.0141	0.0169
Average PRMSE 2010	0.0087	0.0275
Average PRMSE 2014	0.0120	0.0264
Number of Bayesian p-values between 0.1 and 0.9	8 of 36	8 of 32
Number of negative values	0 of 36	0 of 32
Number of values > 1	0 of 36	0 of 32

4.7. Dirichlet-Dirichlet model with a time variant concentration parameter

Figure 4.10 below shows signs of the same issues as the basic Dirichlet-Dirichlet model, where the expected values of the latent variable indicates a volatile behaviour of vote intention that does not corresponds with expectations of dynamics of political opinion or vote intention. That is, the estimations of the latent time series is wiggly in such a way that a large number of people change their vote intention from one day to the other in a unlikely way, even in times where a specific events such as the previously mentioned Juholt scandal have not occurred. This behaviour is similar to when using data pre-processing technique 1 in the basic DLM model in section 4.3. However, the non-smoothness of the expected values of the posterior distributions have not increased from when assuming that the concentration parameter is constant in an obvious way.

The wigglyness of the estimations for the Swedish Democrats are even more apparent towards the end of the time series, when the polling results starts to diverge depending on sampling method. The model once again estimate the latent variables to follow the polls coming from houses not using web panels more closely but the expected values is close to being in-between the two clusters of polling result.

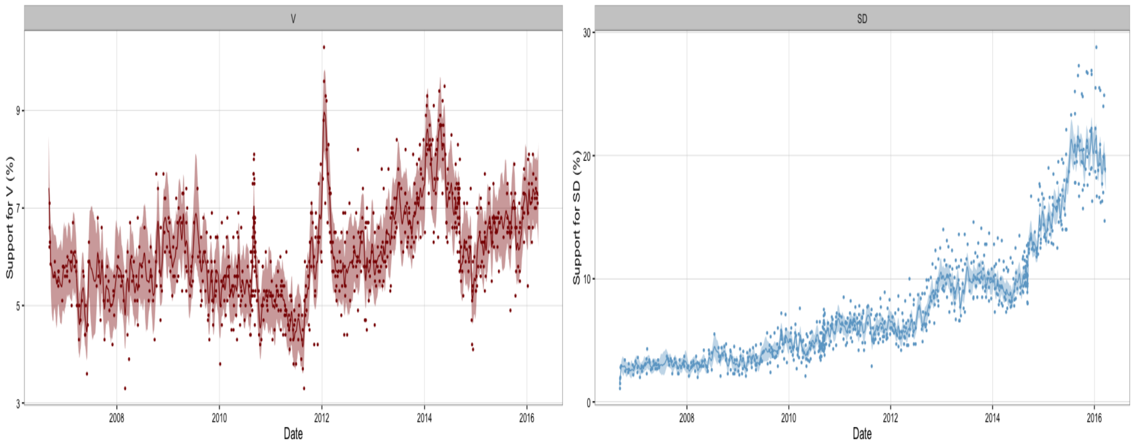

Figure 4.10.: Expected values of the posterior distributions for the latent variable as a line in a Dirichlet-Dirichlet model with a time variant concentration parameter, polling data as points and ribbon for 95% credible bands. V(left) and SD(right).

The non-smoothness of the expected values are perhaps a bit worse than in the basic Dirichlet-Dirichlet model which is explained by the estimations of the concentration parameter. The expected values of the concentration parameter over time can be seen in figure 4.11 along with the posterior distribution of the parameter in appendix B.7 - Figure B.16. It can be seen that when using a time variant concentration parameter yields much lower expected values than in the basic Dirichlet-Dirichlet model when a time invariant concentration parameter was used. Studying the expected values over time it can be seen that it is centered around the mean of the prior for the initial value of the concentration parameter at 10000. Further, the smallest expected value is sligthly lower than 9900 and the highest around 10050, meaning that the concentration parameter does not vary greatly over time.

In Figure 4.11 below it can be seen that the expected values of the concentration parameter decrease at the end of the year 2011. This corresponds with the time of the Juholt scandal, indicating that the vote intention in Sweden indeed where less stable at the time of the scandal which could be explained by voters intending to switch between parties.

A decrease of the concentration parameter can also be seen in 2009, which is consistent with the rapid changes in polls for MP, M, KD and O to which we have not found a historical explanation.

Also, the concentration parameter decrease before the election in 2014 and the expected values of it are noticeably lower at the time of the 2014 election compared to one in 2010. This indicates that state of vote intention was less stable before the election 2014 comapred to 2010, which might be a reason to why all previous models have had more difficulties capturing the true election results in the latest election.

Further, previous work states that Swedes report the intent of changing their vote from one party to another more often between election, indicating that the concentration parameter should decrease between election. This cannot be confirmed by the plot of the expected values of the concentration parameter since there is no real pattern regarding the movement of the stability of vote intention between the three election included in the data.

However, the movement of the expected value of the concentration parameter is limited, which might be an indication that the assumption of the concentration parameter being time invariant is not too problematic.

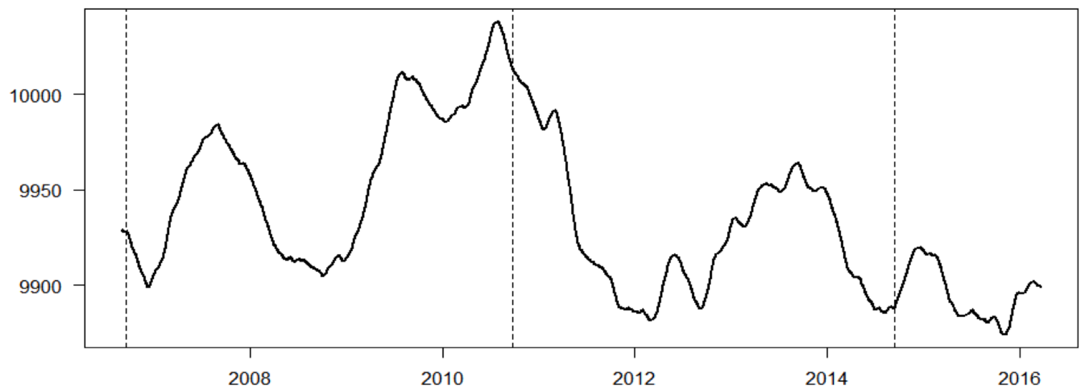

Figure 4.11.: Posterior distribution of concentration parameter.

The prior used when modelling allows for large changes of the concentration parameter over time and should therefore not be restricting the concentration parameter to its initial value. Using another prior for the variance in the random walk of the concentration parameter is used and yields very similar results. That the movement of the concentration parameter is limited is consistent with previous work stating that the Swedish voters in general is not overly dynamic. Looking at the traceplots for the different chains for some of the concentration parameters it can be seen that there is problems with the mixing. The traceplot for one of the chains for one randomly selected concentration parameter can be seen in Figure 4.12 below. It is clear that there the full posterior distribution have not been explored, and that the sampler gets stuck in certain parts of the posterior and showing signs of high autocorrelation between samples. This is a known problem with the random walk Metropolis-Hastings, which is the sampler used for very uncommon posteriors within JAGS. The issues continue after increasing the number of iterations and thinning of the chains.

A reparametrization of the model, or even longer runs of the chains with increased thinning would be needed to obtain independent samples from the posterior. Further attempts where with more aggressive thinning and increased number of iterations could have been made with additional time. The mixing problem of the concentration parameters would effect the accuracy of the inference regarding the other parameters in the model as well.

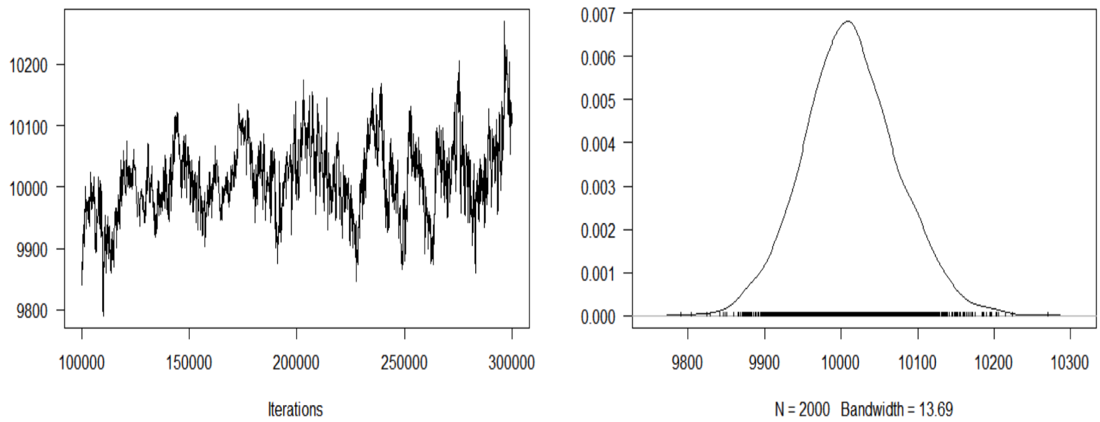

Figure 4.12.: Traceplot for one chain of one randomly selected concentration parameter

In Table 4.15 below it can be seen that the model is equally good as the basic Dirichlet-Dirichlet model in 2010, where the election results for MP and SD are the only ones outside the credible bands. As in all previous models fewer of the true election results are captured within the credible bands in 2014. However, the Liberals are the only party whose credible bands cover the true election result in 2010 but not in 2014.

Thus, this is the best model with regards to similarties between point estimates of the posterior distribution of the latent state on Election Day and the election result. This might have contributed to the variability in the model noticeable in Figure 4.10 where the credible bands cover more of the data points than all previous models, and if comparing the width of the credible bands in Table 4.15 with the ones in Table 4.13, where all credible bands are wider when the concentration parameter is time variant.

MP is once again the party with the highest PRMSEs in 2010, just like in the basic Dirichlet-Dirichlet model, followed by the two major parties and SD. This has been a consistent result in all the models explored in this thesis. The highest PRMSE is recorded for SD in 2014, followed by MP, S and M. Also, all PRMSE are lower when the credible bands cover the election results just as in the basic Dirichlet-Dirichlet model. Furthermore, the fact that the election result does not sum exactly to 100 is due to approximation errors.

Table 4.15.: Point estimates and 95% credible band using data pre-processing technique 3 in a Dirichlet-Dirichlet model with time variant concentration parameter. Bands containing the election results are marked in bold. Election year 2010 and 2014.

Party	Year	Election results	$E(x_{Election})$	95% CB	PRMSE
M	2010	30.1	28.6	**27.0;30.3**	0.0174
	2014	23.3	23.4	**21.2;25.6**	0.0172
L	2010	7.1	7.0	**6.1;8.0**	0.0056
	2014	5.4	7.1	5.9;8.1	0.0067
KD	2010	5.6	5.8	**5.0;6.7**	0.0056
	2014	4.6	5.7	**4.6;6.7**	0.0113
C	2010	6.6	7.3	**6.3;8.2**	0.0090
	2014	6.1	6.8	**5.7;7.8**	0.0100
S	2010	30.7	30.1	**28.4;31.8**	0.0115
	2014	31.0	30.5	**29.3;31.9**	0.0175
MP	2010	7.3	9.2	8.2;10.3	0.0200
	2014	6.9	8.1	7.1;9.6	0.0184
V	2010	5.6	6.0	**5.1;6.9**	0.0066
	2014	5.7	5.9	**4.7;7.3**	0.0058
SD	2010	5.7	4.6	3.9;5.4	0.0115
	2014	12.9	9.1	8.6;10.7	0.0751
O	2010	1.4	1.3	**0.9;1.7**	0.0029
	2014	4.1	3.8	**1.6;4.9**	0.0072

Looking in Table 4.16 it can be seen that the average MAD is slightly higher than for the benchmark model in 2010, while the RMSE is higher in a more obvious way. Both MAD and RMSE is lower than for the benchmark in 2014, which make intuitive sense since more of the true election values are within the credible bands.

The average PRMSE for both years are lower than the benchmark model, but not as low as for the basic Dirichlet-Dirichlet model. This might be due to the an increased variability in the model due to the increased instability of vote intention, reflected in the lower value of the concentration parameter.

The model performs on par with the basic Dirichlet-Dirichlet with regards to the posterior predictive checks, where eight out of 36 values are between 0.1 and 0.9. Thus, this is also a lower proportion compared to the benchmark model. Once again no negative values or values above one is amongst the replicated data due to the nature of the Dirichlet distribution, which can be seen in the appendix B7 - Table B.10.

Table 4.16.: Summary model evaluation measurements Dirichlet-Dirichlet model with time variant concentration parameter.

	Dirichlet-Dirichlet model with time variant concentration parameter	**Benchmark model**
Number of election result covered by credible bands 2010	7 of 9	4 of 8
Number of election result covered by credible bands 2014	6 of 9	1 of 8
MAD 2010	0.0071	0.0070
MAD 2014	0.0106	0.0144
RMSE 2010	0.0093	0.0083
RMSE 2014	0.0151	0.0169
Average PRMSE 2010	0.0100	0.0275
Average PRMSE 2014	0.0188	0.0264
Number of Bayesian p-values between 0.1 and 0.9	8 of 36	8 of 32
Number of negative values	0 of 36	0 of 32
Number of values > 1	0 of 36	0 of 32

Dropping the assumption that stability of vote intention is constant over time is appealing from a theoretical point of view. However, from the results presented in Table 4.16 in can be concluded that using a time variant concentration parameter in the Dirichlet-Dirichlet model does not improve it greatly with regards to many of the evaluation measurements used in this thesis, but perform equally good in most regards. However, it is by far the model for which the 95% central credible bands of the posterior distribution of the latent state at the Election Day captures most of the election results in 2010.

The similarities between the models using a time variant and time invariant concentration parameter might be explained by the fact that the estimated movement of the parameter is limited. The difference between the excepted values through time is only around 150. This could indicate that political scandals leading to large changes in the poll only corresponds to a small movement in actual vote intention, signaling that some of the differences in polls might come from house effects.

However, by letting the concentration parameter vary over time the analysis of the output of the model can deepen in a way that relates to previous works from other domains, but does not seem to be necessary when assuming both the latent state and the observed data to be Dirichlet distributed. The increased complexity of the model creates issues with sampling using the JAGS software, making it desirable to use other programs for sampling in future work.

5. Discussion

The following discussion has been divided into sections corresponding with the sub questions presented in section 1.3.

i. Dealing with periodically collected data From the results when using the same basic dynamic linear model dealing with the periodically collected data in different ways, it can be concluded that the model performs very differently. This is an interesting result since previous work largely uses only one data pre-processing technique without commenting further on the effects of the choice, and simulations studies or comparisons such as performed in this thesis have as far as we know never been done in the domain. Seeing the observed data as the average of the latent states during the collection period is appealing from a theoretical viewpoint since it captures the fact that the true observed data at each collection day is unknown to everyone except the polling houses themselves. Comparing this approach to the other data pre-processing techniques used in this thesis it appears to be a promising novel way of handling periodically collected data when creating a Bayesian poll of polls. It is also encouraging that the data pre-processing techniques yields similar results regardless of the assumed distribution, since the same technique seems favourable using both Gaussian and non-Gaussian models.

The data pre-processing technique assuming that the result from the poll is equal over the entire collection period seems to overestimate the certainty of the polling results, decreasing the variability in such a way that the model loses its ability to capture extreme values in the measurement equation. This assumption also restricts the models' ability to capture plausible large changes in the latent state. Letting one time point in the middle of the data collection period represent the whole poll yields estimations of the latent states that are heavily affected by each new poll, leading to an unconvincingly behaviour of the states. However, one advantage of using data pre-processing technique 2, where each poll is divided into multiple smaller polls and thus increasing the size of the data, is the fast convergence. The increased speed indicates that the modelling could benefit from more data. Pooling of data from several countries with a similar political arena, such as the Scandinavian countries, to increase the data size and investigate how this affects the modelling is a possible prolongation for future work. Another solution could be aggregating data pooling polling results from the same week, or month, and using the average leading to less time points without any data. The need for deeper consideration of how to deal with periodically collected data and documentation of the effect of these choices when using dynamic linear models is desirable in future works in the domain.

ii. Distribution assumptions Discarding the widely used normality assumption, instead using the multivariate Dirichlet distribution for both the measurement and the dynamic equation, yields models for which the three out of eleven evaluation measures are on par with the univariate Gaussian models and outperform them in five. The Dirichlet-Dirichlet models are previously overlooked even if it theoretically is suitable for the domain. One can speculate that the appearance of the latent states would be more consistent with expectation if the concentration parameter was estimated to a higher value, even if there is no previous research to compare the results with. The prior used covers much higher values and should therefore not be restricting the estimated posterior of the concentration parameter. Both more and less informative priors were also investigated with poorer or equal results. Further, the results show how the size of the parties affects the estimations, where the expected value of the latent variable for the larger parties has a more appealing appearance. In Stoltenberg (2013) the concentration parameter in a Multinomial-Dirichlet model was set to 1000, using weekly average of observed data, with smooth estimations of the latent states as a result. However, when experimenting with different priors for the concentration parameter it is obvious that 1000 is a much to low value when modelling using data pre-processing technique 3. This is a further indication that the data pre-processing technique will have an affect on other modelling decision and the results.

When the concentration parameter in the Dirichlet model are modelled as time variant the movement of the concentration parameter is still somewhat restricted, where the expected value of the parameter only differed with around 150 between the lowest and highest value. Problems with mixing in the simulations for the concentration parameter are detected which could be a probable explanation to the issues with the model. The issues with mixing of these parameters could also have affected the inference of the other parameters, and thus the other results found when modelling. The fact that the expected value of the concentration parameter do not vary greatly through time might indicate that the assumption of a constant concentration parameter is valid, confirming previous studies of the dynamics of Swedish voters. However, the movement of the concentration parameter estimated in the model does not confirm previous work suggesting that the stability of vote intention is lower between election than in election campaigns. The results in this thesis rather implies that vote intention becomes more dynamic just before election, a result that was especially clear before the 2014 election, which could explain the many failed election predictions made before the election.

Comparing the results from a basic multivariate and univariate model indicate that assuming independence between parties in a multi-party system does not create obvious problems making inference of the latent states nor yields implausible solutions when applied on Swedish polling data. However, the multivariate case has a clear advantage in terms of interpretations and theoretical connection to the nature of the problem dealt with in poll of polls. The concentration parameter in the Dirchlet-Dirichlet model yield further possibilities to combine knowledge regarding vote intention from other domains into the model, such as the stability of the parameter. This is an area which, to our knowledge, is still uncharted within the domain of poll of polls modelling. However, using Dirichlet-Dirichlet models yields none standard posterior distributions, increasing computational time the making the already slow sampler less attractive to use. Thus, implementing the models using a different program using other MCMC methods such as the Hamiltonian MCMC used in the modelling language STAN could be useful for future work. To compare the complexity using different sampling methods, and to implement the models in another program is a contribution to be made in future work. However, working with the JAGS is a simple and well working option that already is an established tool in the domain.

The difficulties with the Dirichlet-Dirichlet models are rather unexpected since it theoretically is very well suited for the nature of the problem, and thus it is still relevant to extend the work of this model in the future. Even if this first attempt of multivariate models assuming both the latent and the observed variable are Dirichlet distributed do not outperform the univariate Gaussian models in an obvious way, there is still many different models to investigate in future work on poll of polls models using dynamic linear models in multi-party system. Attempts of using a Multinomial-Dirichlet model was also done, but with similar results to the Dirichlet-Dirichlet model it was not presented in the thesis but could be more deeply investigated in future work. Another such interesting model would be to assume the latent variable to follow a multivariate normal distribution and the observed data to follow a Dirichlet distribution. This novel approach using a multivariate normal distribution for the latent states would allow analysis of the correlations between the different political parties, which can be compared to previous works on the dynamics of Swedish voters. An additional model for future work could be trying to implement a Pólya-Gamma Augmentation, which allow for dependence in multinomial variables and thus theoretically a very interesting approach. Further, the results with an univariate Gaussian model with house effects included still have issues capturing the more extreme values in the observed data. In future work it could be interesting to see if using a fatter tailed distribution, such as the student-t, yields better results.

iii. Model components The results in this thesis also indicate that house effects could be included when creating a poll of polls model in a multiparty system with univariate Gaussian models. However, the model does not outperform the models without house effects in an overwhelming way. This can be due to the fact that house effects are indeed not present in the polls or that the house effects used in this thesis do not capture the house effects in a satisfying way, where the second explanation is the more probable. As it have been shown there is multiple ways of including house effects, which have different interpretations and results. Studies using different house effects with comparative result are conspicuous by its absence within the domain. By trying out different house effects, both time variant and time invariant, it is apparent that they yield rather similar conclusions and equally performing models in many of the evaluation measurements. By letting the house effects affect the variance in the measurement equation multiplicatively it can be interpreted as how the efficiency of the sampling is different from what would be expected from simple random sampling. The results in this thesis indicate that adding the house effects multiplicative on the variance is a more attractive way of modelling house effects than additive on the mean which is a common approach in current research. The interpretation in terms of capturing how the efficiency compared to a random sample is also intuitive. Similar results are found when house effects are added additive to the variance. However, the increased variability using the additive house effect is expected due to the construction of the model where the prior only allow positive values. An improvement in future models would be investigating the performance of an additive house effect on the variance where the house effect can decrease the variability in the measurement equation as well.

Using multiplicative time variant house effects on the variance enables studies of how the uncertainty in the polling has changed over time. This information could be especially interesting for polling houses when evaluating the effects of methodology changes. Further, a vague normal prior for the additive house effects is used in this thesis. This vague prior is somewhat contradictory to restricting the sum of the house effects for the parties and houses to zero, since this implies that the houses in general have no bias. Assuming that houses in general have no bias is not consistent with previous work regarding Swedish polling data, but has been suggested in studies in other countries. Relaxing this restriction does however lead to poorly working models due to issues with identification in between elections, but it is desirable to find a way of relaxing this constraint in future work. Another possible continuation for future work is to use a prior for the time variant house effects that manage to capture the plausible dual behaviour in house effects, the abrupt movement of the effect that might occur after a methodology change and the slow moving process consistent with constant issues such as low response rates or untruthful answers due to the sensitivity of the question. An additional interesting extension for future work regarding house effects could be using a prior allowing for shrinkage, such as a horseshoe prior, investigating the presence of house effects even further. Initial attempts with house effects that depend on the number of days since the latest election yielded very small house effects, which could indicate a necessity of shrinkage.

The modelling in this thesis has also focused on the measurement equation, especially if and how house effects should be incorporated in a poll of polls model in a multi-party setting. More work can be done in the dynamic equation, where knowledge regarding voting behavior of sympathizers of the different parties can be incorporated, such as a variable capturing the loss in vote intention for the major governing party before election. These kinds of components require more knowledge of the dynamics of the political opinion in the political context in which the model is adapted than the more universally usable house effects that is a consequence of survey methodology rather than country specific, making them more difficult to generalize.

Trials with house effects in the Dirichlet-Dirichlet model have also been done by using additive house effects on the proportions and multiplicative house effects on the concentration parameter when a time variant concentration parameter has been used, both so far with poor results. The failed attempts might be a result of estimations of the concentration parameter and the house effects cancel each other out. It is therefore interesting to study house effects in a Dirichlet-Dirichlet model that assume the concentration parameter to be constant. Initial attempts of using time variants house effects, have also been investigated in the model but with another poor result, and are thus no included in this thesis, but can be interesting for future work.

Further, none of the models investigated in this thesis manage to show increased variability when the variance in the observed data increases. The most obvious example of this is the Swedish Democrats where the polling data diverge, with a high variance between polls using web panel and other methods. A continued challenge for future work is to capture this behaviour. One approach could be including a variance component that reflects the degree of data divergence, by multiplying the variance in the measurement equation with the difference between the highest and lowest recorded vote proportion within the latest month. This would yield an increased uncertainty of the estimations of the latent states when the variance is high amongst observed data.

iv. Evaluation When evaluating the models using both traditional measurement for evaluating Bayesian models in general and poll of poll models have been used, as well as the Posterior RMSE. Comparing simple test quantities, such as the minimum, maximum and mean value, in a posterior predictive check indicates that replicated data from the posterior predictive distribution generated from the same distribution as the observed data. That is, the model generating the data does not seem to be amongst the ones tried in this thesis. The models showing a very volatile behaviour of the latent state, different from would be assumed taking into account knowledge of voting behaviour from political and behavioural science, will show better results in terms of these posterior predictive measurements. This might be an indication of that the traditional evaluation measures for the posterior predictive distribution favours models with high variability, or simply that these models are more similar to the one generating the data since the data is noisy.

The posterior predictive checks also indicate that the benchmark model is more similar to the one generating the observed data than all other models presented in this thesis. This is due to the fact that it captures one feature, the mean, of the distribution very well. Thus, it would be interesting trying to find more fitting test quantities for this specific domain. Increasing the statistical testing of models from what is common practice today would decrease the dependency on background knowledge of the domain, making it attractive to a wider range of scientist, as well as reducing the subjective evaluation of which model is appropiate to use. Using PRMSE captures both the variance in the posterior distribution and the deviation between the true election results and the simulated point estimates. In most univariate models there is one instance where the PRMSE is lower when the point estimate is not captured by the 95% central credible bands, while this behaviour is non existing in the Dirichlet-Dirichlet models. Further, the use of the binomial variance leads to that the variance in the measurement equation is larger for parties with vote intention proportion close to 0.5. This might be the reason to why the PRMSE measurements for the two major parties M and S consistently are amongst the highest, possible due to high variance amongst the draws from the posterior distribution. Further, the very traditional MAD and RMSE measurement is a good complement to the PRMSE since a model that have a high variance amongst the draws from the posterior still have point estimates close to the election results. Even these standard evaluation measurements are relatively unused in the domain.

The benchmark model is today published monthly by commission of a public service broadcaster, in this thesis it has been shown that the model simply captures the mean of the polling average of polls conducted within the same month, where the largest drawback of the model is the lack of interpretability. One makes assumptions that vote intention is centred around the polling average, which is as to say in advance that the polling houses on average indeed capture the true vote intention in a country. Thus, not taking into account the issues with house effects found in previous research as well as in this thesis. The results for the benchmark model indicate that without any previous knowledge about the domain one can still produce a model that is relatively well-behaved. It might also be an indication that one should aggregate polls during a specific time period, for example during a week or month, to have shorter periods with missing data. However, by using dynamic linear models higher interpretability and better results with regards to most of the evaluation measured used in this thesis are obtained.

6. Conclusions

The method of dealing with periodically collected data in the dynamic linear model will heavily influence the outcome of the modelling, and should be considered more in poll of polls model. By pre-procesing the the data and introducing it into the models in different ways one can achieve desirable results, but not all of them might correspond well with the periodic way data is collected.

Further, the univariate Gaussian models that have dominated the domain seem to hold a similar capacity as models assuming that the observed and the latent variable is generated from a Dirichlet distribution. However, by using models that assume that the political parties are dependent on each other corresponds better to the nature of the problem at hand and there is still many different combinations of distributions of the observed variables and the latent states to consider for future work of obtaining poll of polls models in multi-party systems.

The results in this thesis suggest that when creating a bayesian poll of polls model for a multi-party system, using univariate Gaussian dynamic linear models, house effects could be included. Multiplicative house effects on the variance seem the most promising, under the restrictions utilised in this thesis, amongst the ones tested. This house effect have an natural interpretation fitting the problem and problems with identification seem to be avoided. Further works of components that could affect the dynamic part of the models can still be done.

In previous work evaluation has often been done visually and by comparing point estimates at Election Day. The evaluation measurements used in this thesis were an attempt to apply traditional evaluation measurements for Bayesian modelling to the domain, where the different measurements mostly have favoured the same models. However, the evaluation measurements indicate that none of the models seem to be the one generating the data.

A. Software

A.1. Markov Chain Monte Carlo in JAGS

The main interest within the Bayesian framework is to make inference regarding the posterior distribution. The general method for simulating from the posterior distribution is using Markov Chain Monte Carlo (MCMC) simulations, where parameter are simulated from distributions approximate to the target posterior distribution and then corrected the drawn valued to better correspond to the ones in the target distribution. The most commonly used way of implementing the DLMs in research within political polling have been using JAGS (Just Another Gibbs Sampler), which mainly employs the Gibbs sampler for the MCMC algorithm. The Gibbs sampler is adapted for multidimensional target distributions, and works by producing a Markov chain whose stationary distribution is the same as the posterior distribution of the hidden state, the target distribution. The Gibbs sampler divides up complex problems into series of easier problems, by sampling sequentially from univariate conditional distributions (Givens and Hoeting, 2013). That is, by obtaining the full conditional distribution of each parameter on the other parameters the joint distribution of these parameters are obtained.

JAGS sets up the problem as a Directed Acyclic Graph (DAG) including stochastic and non-stochastic nodes. The posterior conditional density of a node in the DAG of the model is found by the Markov blanket of a node, which joint distribution contains sufficient knowledge to calculate the posterior distribution of said node; $f(\theta_j|\mathcal{G} \setminus \theta_j) \propto f(\theta_j|parent\,[\theta_j]) \prod_{\tau \in children[\theta_j]} f(\tau|parents[\tau])$.Where $G \setminus \theta_j$ is notation from set theory and refers to nodes in the relative complement. That is, the conditional distribution for a node is the product of the conditional distributions on its parent nodes and its children nodes (Stoltenberg, 2013). Looking at the figure of the DAG for the basic DLM model in (3.9) below, one can see that the posterior conditional distribution would be proportional to the following product $f(x_{kt}|x_{kt-1}, \sigma^2_{w_{kt}}) f(x_{kt+1}|x_{kt}, \sigma^2_{w_k}) f(y_{kt}|x_{kt}, \sigma^2_{v_{kt}})$. In equation 3.6 one can see that all of these distributions are assumed to be Gaussian and therefore the posterior will be Gaussian as well. In non-conjugate cases adaptive rehection sampling is used when if the full-conditional density is log-concave otherwise slice sampling or random walk Metropolis-Hasting will be used instead of the Gibbs sampler. The documentation of how the choice between slice sampling and random walk Metropolis-Hasting is scarce, however both samplers should theoretically converge even if one might be slower than the other to reach convergence.

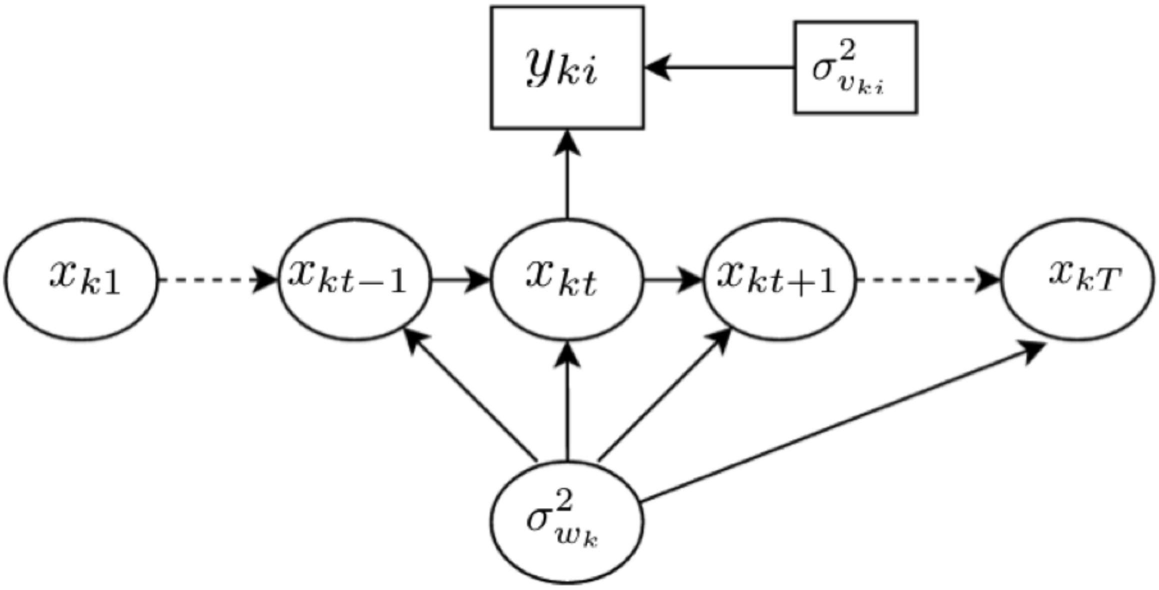

Figure A.1.: Directed Acyclic Graph for the basic DLM model as created by JAGS to obtain posterior distributions of each node.

B. Plots and figures

B.1. Plots and tables of results for section 4.2 – the benchmark model

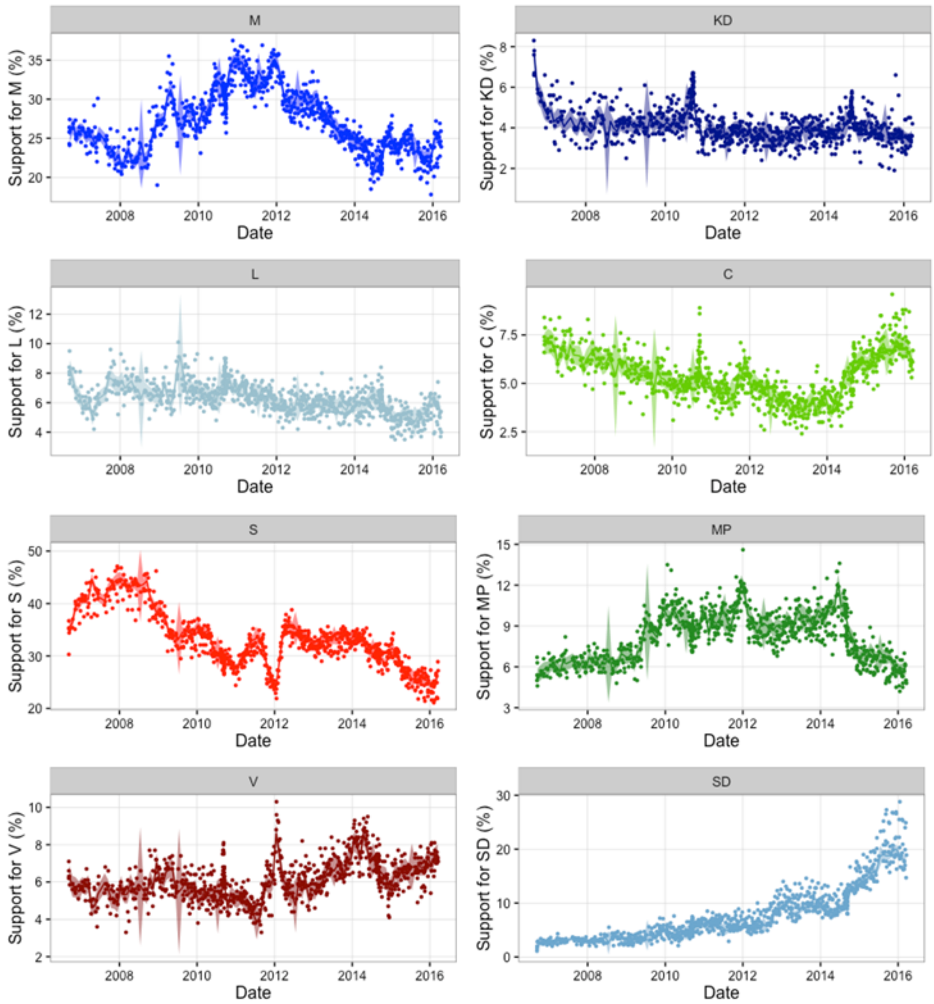

Figure B.1.: Expected value of the posterior distributions for the latent variable for all political parties using benchmark model, with 95% credible bands.

Table B.1.: The probability of the test quantities minimum, maximum, mean and variance in the replicated data are greater than in the observed data. The percentage of negative values and vote intention proportions larger than 1 to the left.

	Benchmark model					
	Min	Max	Mean	Variance	Negative values	Above one
M	1.000	0.000	0.526	0.128	0.000	0.000
L	0.997	0.093	0.712	0.000	0.000	0.000
KD	0.999	0.013	0.006	0.000	0.000	0.000
C	0.972	0.170	0.594	0.016	0.000	0.000
S	0.979	0.004	1.000	0.160	0.000	0.000
MP	0.991	0.000	0.400	0.192	0.000	0.000
V	0.989	0.003	0.001	0.000	0.000	0.000
SD	1.000	0.000	0.000	0.000	0.000	0.000

B.2. Plots and tables of results for section 4.3 – basic dynamic linear model

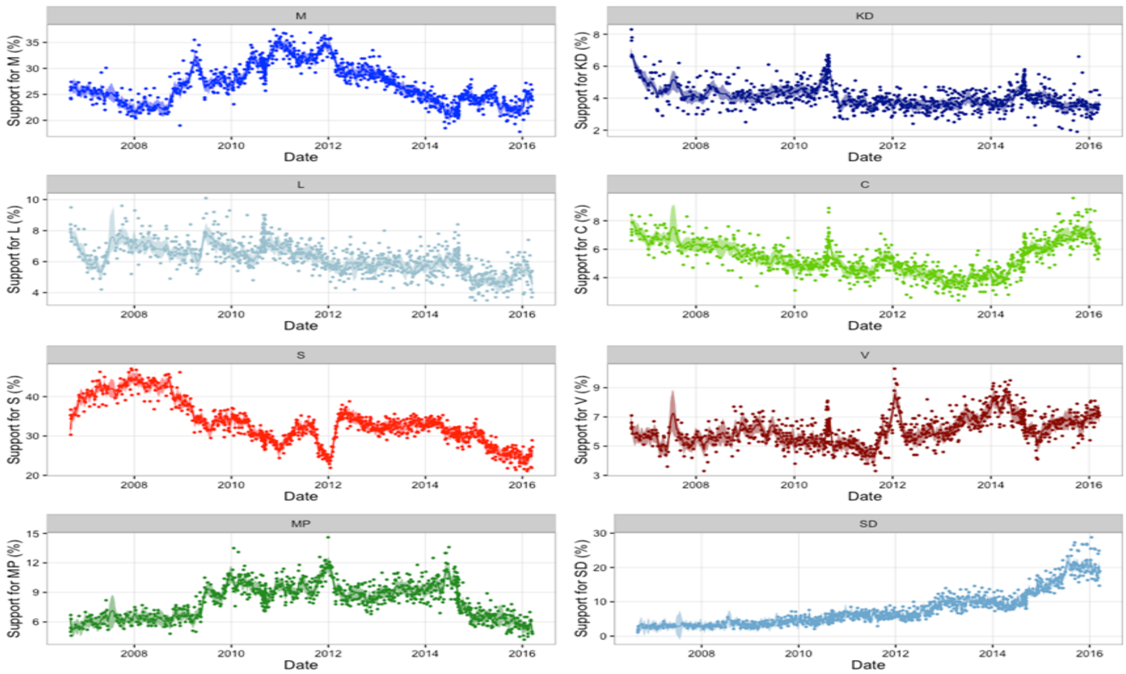

Figure B.2.: Expected value of the posterior distributions for the latent variable for all political party using data pre-processing technique 1, with 95% credible bands.

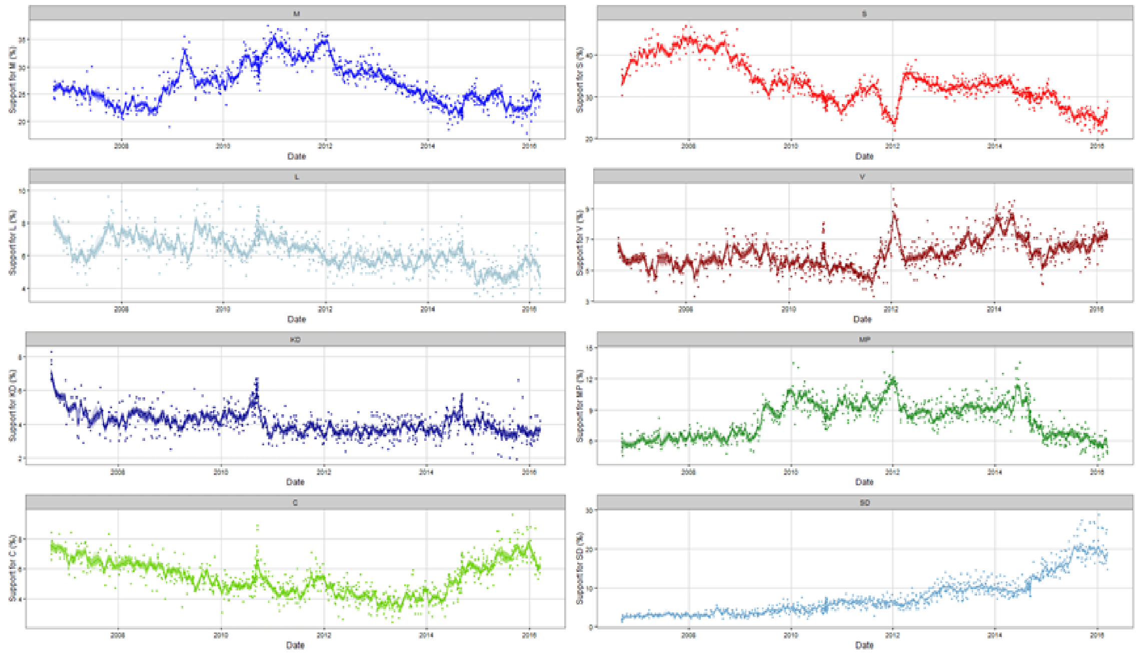

Figure B.3.: Expected value of the posterior distributions for the latent variable for all political party using data pre-processing technique 2, with 95% credible bands.

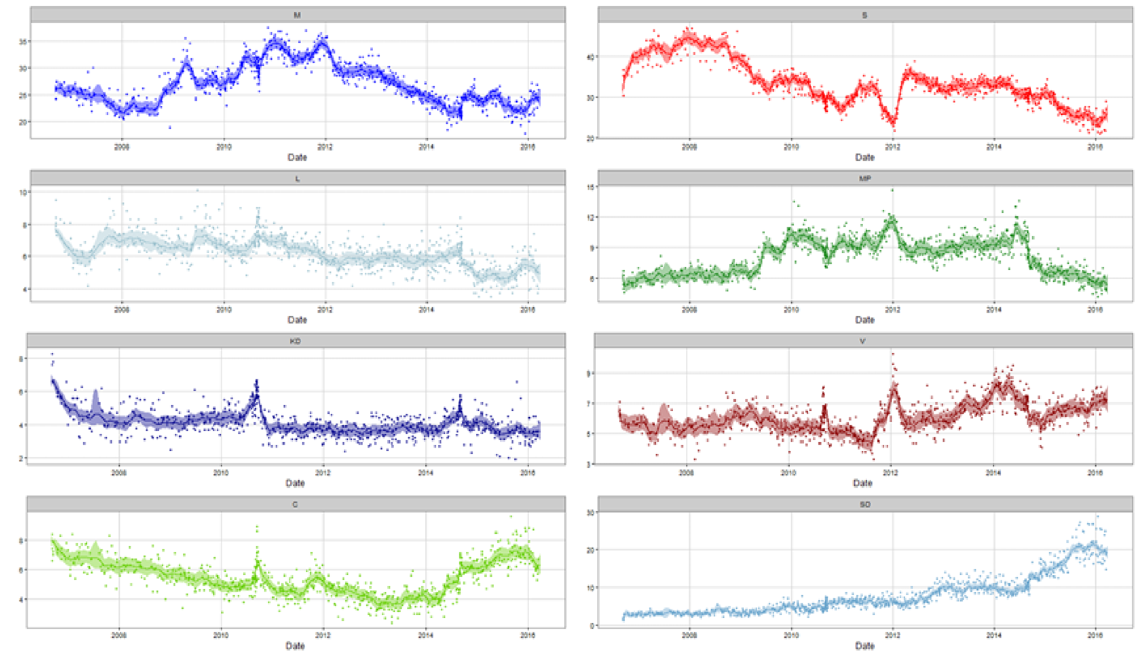

Figure B.4.: Expected value of the posterior distributions for the latent variable for all political party using data pre-processing technique 3, with 95% credible bands.

Table B.2.: Point estimates and 95% credible bands, using data pre-processing technique 1 and 2. Bands containing the election results are marked with bold. Election year 2010.

		Basic DLM			**Basic DLM**		
		Data pre-processing technique 1			**Data pre-processing technique 2**		
	Election results	$E(x_{Election})$	95% CB	RMSE	$E(x_{Election})$	95% CB	RMSE
M	30.1	28.8	**27.1;30.4**	0.0519	28.9	28.1;29.8	0.0557
L	7.1	7.3	**6.7;7.9**	0.0063	7.4	**7.2;7.6**	0.0100
KD	5.6	5.9	**5.5;6.3**	0.0134	5.9	5.7;6.1	0.0100
C	6.6	6.7	**6.2;7.2**	0.0103	5.7	5.6;5.9	0.0100
S	30.7	29.2	**26.5;31.9**	0.0816	28.9	28.4;29.4	0.0755
MP	7.3	8.8	8.1;9.5	0.0166	8.8	8.6;9.0	0.0224
V	5.6	6.1	**5.5;6.6**	0.0135	6.3	6.1;6.5	0.0141
SD	5.7	5	**4.0;5.9**	0.0238	4.5	4.3;4.7	0.0173

Table B.3.: Point estimates and 95% credible band, using data pre-processing technique 1 and 2. Bands containing the election results are marked with bold. Election year 2014.

		Basic DLM			**Basic DLM**		
		Data pre-processing technique 1			**Data pre-processing technique 2**		
	Election results	$E(x_{Election})$	95% CB	RMSE	$E(x_{Election})$	95% CB	MSE
M	23.3	21.9	20.9;22.9	0.0485	22.3	21.8;22.9	0.0548
L	5.4	6.5	6.1;6.8	0.0093	6.4	6.1;6.7	0.0141
KD	4.6	6	4.8;5.4	0.0173	5.1	4.9;5.3	0.0100
C	6.1	5.1	**5.6;6.4**	0.0168	6.1	**5.9;6.3**	0.0141
S	31.0	30.3	**28.8;31.9**	0.0608	30.0	29.6;30.4	0.0548
MP	6.9	8.8	8.2;9.5	0.0138	8.8	8.6;9.1	0.0100
V	5.7	6.7	6.1;7.2	0.0188	6.8	6.6;6.97	0.0316
SD	12.9	9.9	8.5;11.2	0.0271	10.2	9.9;10.4	0.0707

Table B.4.: The probability of the test quantities minimum, maximum, mean and variance in the replicated data are greater than in the observed data using the basic DLM and data pre-processing technique1, 2 and 3. The percentage of negative values and vote intention proportion larger than 1 to the left.

	Data pre-processing technique 1					
	Min	Max	Mean	Var	Negative values	Above one
M	0.912	0.310	0.247	0.001	0.000	0.000
L	0.999	0.000	0.001	0.000	0.000	0.000
KD	1.000	0.000	0.000	0.000	0.000	0.000
C	1.000	0.000	0.000	0.000	0.000	0.000
S	1.000	0.001	0.221	0.000	0.000	0.000
MP	1.000	0.481	0.000	0.000	0.000	0.000
V	0.995	0.000	0.001	0.000	0.000	0.000
SD	0.593	0.000	0.000	0.000	0.000	0.000

	Data pre-processing technique 2					
	Min	Max	Mean	Var	Negative values	Above one
M	0.000	0.001	0.000	0.000	0.000	0.000
L	0.000	0.000	0.000	0.000	0.000	0.000
KD	0.000	0.000	0.000	0.000	0.000	0.000
C	0.000	0.000	0.000	0.000	0.000	0.000
S	0.000	0.000	0.000	0.000	0.000	0.000
MP	0.000	0.000	0.000	0.000	0.000	0.000
V	0.000	0.000	0.000	0.000	0.000	0.000
SD	0.000	0.000	0.000	0.000	0.000	0.000

	Data pre-processing technique 3					
	Min	Max	Mean	Var	Negative values	Above one
M	1.000	1.000	1.000	1.000	0.000	0.000
L	1.000	1.000	1.000	0.120	0.000	0.000
KD	1.000	1.000	1.000	1.000	0.000	0.000
C	1.000	1.000	1.000	1.000	0.000	0.000
S	1.000	1.000	1.000	1.000	0.000	0.000
MP	1.000	0.996	1.000	0.025	0.000	0.000
V	1.000	1.000	1.000	0.824	0.000	0.000
SD	1.000	0.000	1.000	1.000	0.000	0.000

B.3. Plots and tables of results for section 4.4 – dynamic linear model with time invariant house effects

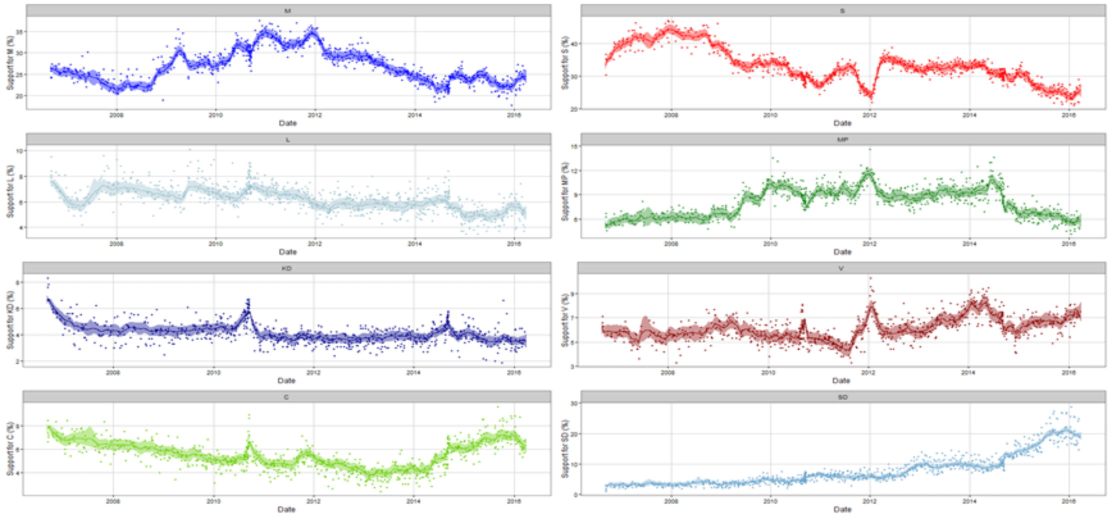

Figure B.5.: Expected value of the posterior distributions for the latent variable for all political party using additive house effects on the mean, with 95% credible bands.

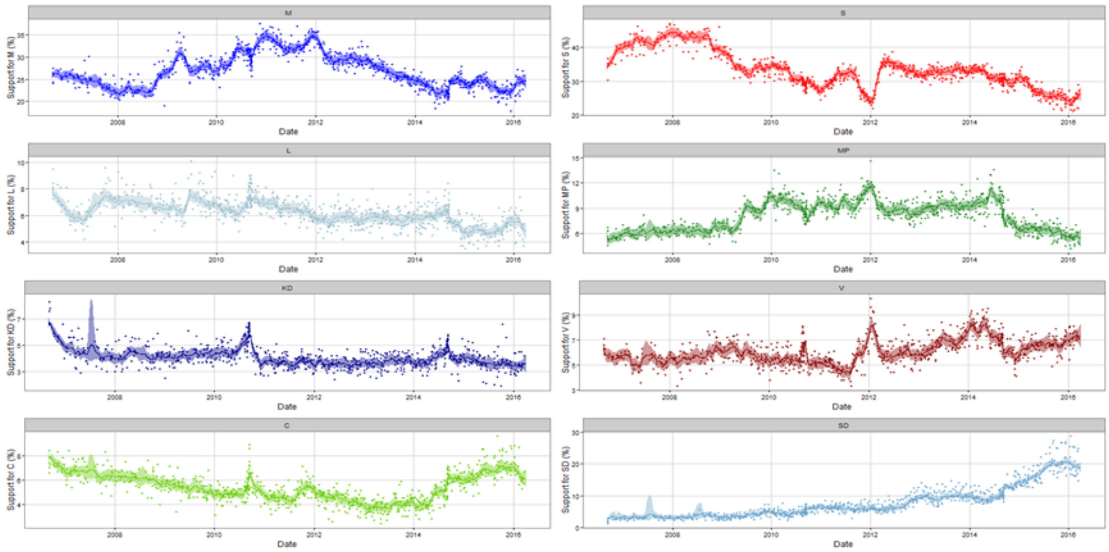

Figure B.6.: Expected value of the posterior distributions for the latent variable for all political party using multiplicative house effects on the variance, with 95% credible bands.

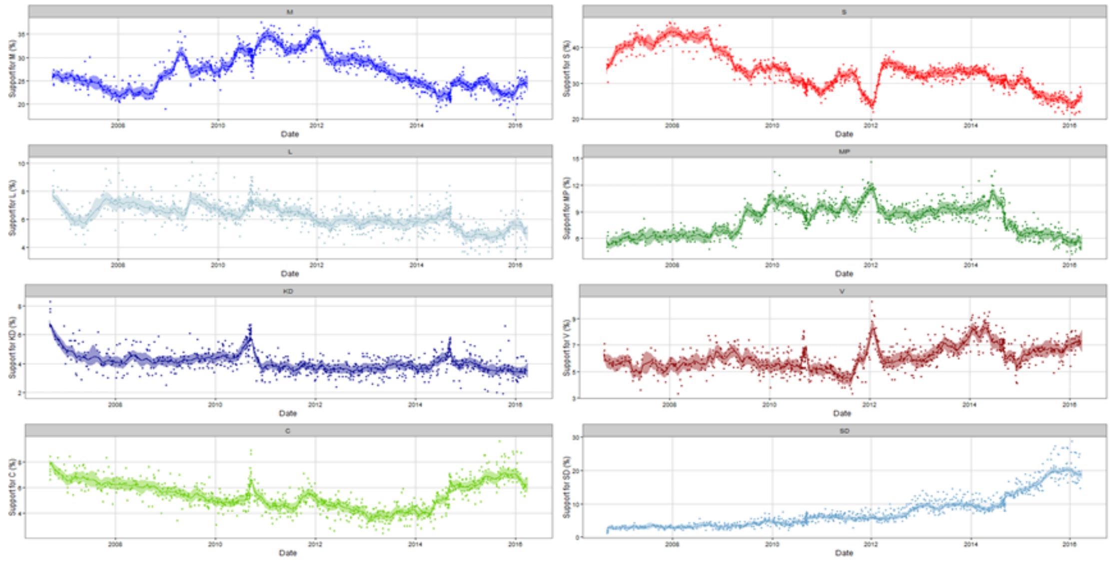

Figure B.7.: Expected value of the posterior distributions for the latent variable for all political party using additive house effects on the variance, with 95% credible bands.

Table B.5.: The probability of the test quantities minimum, maximum, mean and variance in the replicated data are greater than in the observed data using DLM with three different time variant house effects. The percentage of negative values and vote intention proportion larger than 1 to the left.

	Additive on mean					
	Min	Max	Mean	Var	Negative values	Above one
M	1.000	1.000	1.000	1.000	0.000	0.000
L	1.000	1.000	1.000	0.015	0.000	0.000
KD	1.000	1.000	1.000	1.000	0.000	0.000
C	1.000	1.000	1.000	0.996	0.000	0.000
S	1.000	1.000	1.000	1.000	0.000	0.000
MP	1.000	0.998	1.000	0.001	0.000	0.000
V	1.000	1.000	1.000	0.427	0.000	0.000
SD	1.000	0.000	1.000	1.000	0.000	0.000

	Multiplicative on variance					
	Min	Max	Mean	Var	Negative values	Above one
M	1.000	1.000	1.000	1.000	0.000	0.000
L	1.000	1.000	1.000	0.274	0.000	0.000
KD	1.000	1.000	1.000	1.000	0.000	0.000
C	1.000	1.000	1.000	1.000	0.000	0.000
S	1.000	1.000	1.000	1.000	0.000	0.000
MP	1.000	0.997	1.000	0.009	0.000	0.000
V	1.000	1.000	1.000	0.756	0.000	0.000
SD	1.000	0.009	1.000	1.000	0.000	0.000

	Additive on variance					
	Min	Max	Mean	Var	Negative values	Above one
M	1.000	1.000	1.000	1.000	0.000	0.000
L	1.000	1.000	1.000	0.160	0.000	0.000
KD	1.000	1.000	1.000	1.000	0.000	0.000
C	1.000	1.000	1.000	1.000	0.000	0.000
S	1.000	1.000	1.000	1.000	0.000	0.000
MP	1.000	0.998	1.000	0.002	0.000	0.000
V	1.000	1.000	1.000	0.821	0.000	0.000
SD	1.000	0.000	1.000	1.000	0.000	0.000

Table B.6.: Expected value of the posterior distributions for house effects using three different time invariant house effects in a dynamic linear model.

	Additive on mean							
	M	L	KD	C	S	MP	V	SD
Demoskop	0.008	0.001	-0.004	-0.004	0.002	0.003	-0.002	-0.004
Inizio	-0.002	-0.007	0.006	0.010	0.005	-0.01	-0.005	0.003
Ipsos	0.007	0.001	-0.002	-0.002	0.008	-0.001	-0.002	-0.008
Novus	0.001	0.000	-0.001	-0.002	0.004	0.000	0.000	-0.002
SCB	-0.004	-0.004	0.000	-0.001	0.023	-0.004	-0.005	-0.006
Sentio	-0.004	-0.003	-0.001	-0.004	-0.002	-0.004	0.003	0.015
Sifo	0.000	0.002	-0.001	-0.001	0.002	0.004	-0.002	-0.004
Skop	0.011	0.005	0.000	0.001	-0.008	0.002	-0.002	-0.008
SVT	-0.004	0.004	0.001	0.006	-0.016	0.012	0.010	-0.014
United Minds	-0.012	0.003	0.002	-0.002	-0.008	0.003	0.004	0.012
YouGov	-0.001	-0.002	0.000	-0.002	-0.01	-0.005	0.002	0.018

	Multiplicative on variance							
	M	L	KD	C	S	MP	V	SD
Demoskop	1.01	0.23	0.26	0.31	0.1	0.49	0.41	0.66
Inizio	1.05	1.77	3.26	3.37	0.33	1.39	0.77	0.63
Ipsos	0.74	0.46	0.22	0.23	0.58	0.58	0.52	3.68
Novus	0.12	0.13	0.25	0.29	0.24	0.12	0.2	1.11
SCB	1.43	1.14	0.28	0.54	6.51	2.74	1.00	2.47
Sentio	1.35	0.67	0.79	0.74	1.53	0.45	0.2	8.64
Sifo	0.30	0.14	0.23	0.10	0.36	1.03	0.37	1.87
Skop	0.25	0.39	0.69	0.38	1.64	0.10	0.19	1.93
SVT	2.18	1.19	0.63	1.77	7.45	4.14	5.34	10.11
United Minds	1.24	0.16	0.16	0.48	0.74	0.27	0.27	3.37
YouGov	0.48	0.35	0.3	0.62	1.34	0.87	0.16	8.22

	Additive on variance							
	M	L	KD	C	S	MP	V	SD
Demoskop	1.11	0.40	0.11	0.10	0.10	0.68	0.21	0.23
Inizio	1.04	1.89	3.15	3.30	0.33	1.26	0.69	1.08
Ipsos	0.78	0.59	0.18	0.15	0.63	0.62	0.40	2.94
Novus	0.08	0.10	0.12	0.24	0.23	0.08	0.21	0.70
SCB	1.39	1.22	0.24	0.53	6.38	2.70	1.09	1.34
Sentio	1.09	0.48	0.35	0.38	1.56	0.29	0.26	7.64
Sifo	0.31	0.24	0.22	0.12	0.36	1.20	0.28	1.23
Skop	0.29	0.57	0.80	0.52	1.53	0.10	0.15	1.26
SVT	2.13	1.26	0.65	1.81	7.47	4.26	5.17	11.23
United Minds	1.21	0.20	0.18	0.62	0.67	0.38	0.33	3.39
YouGov	0.41	0.20	0.26	0.56	1.31	0.63	0.29	7.73

B.4. Plots and tables of results for section 4.5 – time variant house effects

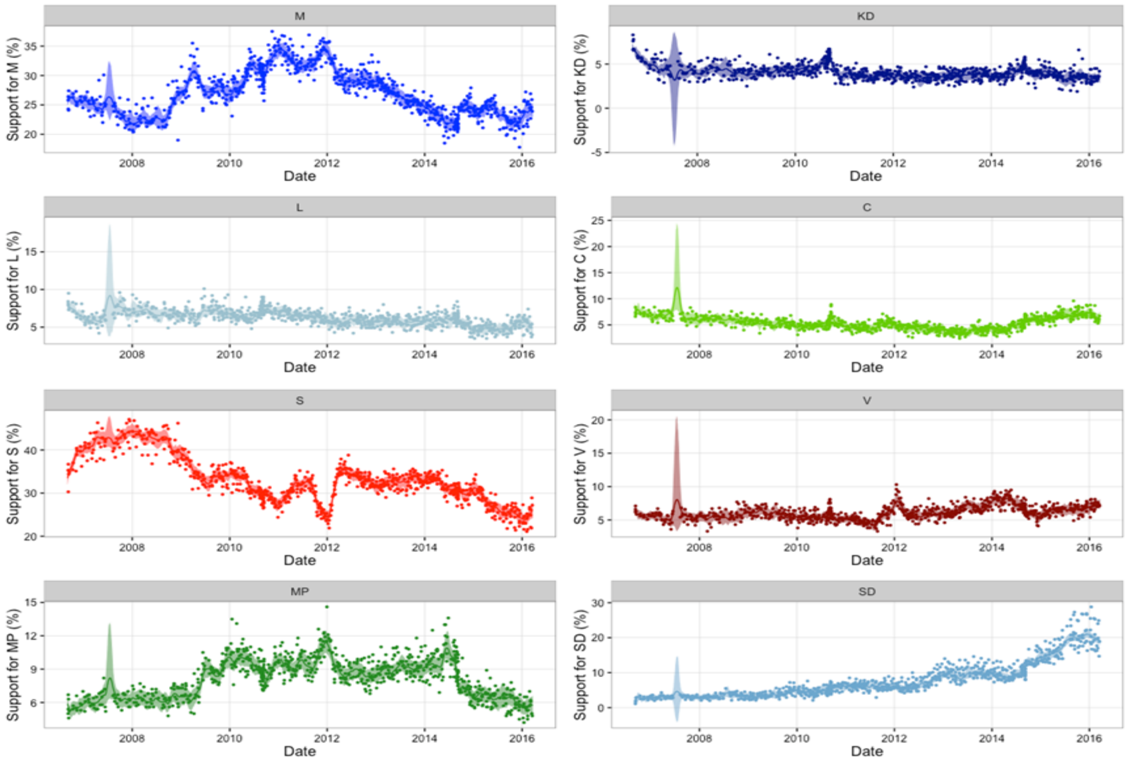

Figure B.8.: Expected value of the posterior distributions for the latent variable for all political party using time variant house effects, with 95% credible bands.

Table B.7.: The probability of the test quantities minimum, maximum, mean and variance in the replicated data are greater than in the observed data using DLM with time variant house effects. The percentage of negative values and vote intention proportion larger than 1 to the left.

	Time variant house effects multiplicative on the mean					
	Min	Max	Mean	Var	Negative values	Above one
M	1.000	1.000	1.000	1.000	0.000	0.000
L	0.977	1.000	1.000	0.933	0.000	0.000
KD	0.897	1.000	1.000	1.000	0.000	0.000
C	0.952	1.000	1.000	1.000	0.000	0.000
S	1.000	1.000	1.000	1.000	0.000	0.000
MP	1.000	1.000	1.000	0.285	0.000	0.000
V	0.968	1.000	1.000	1.000	0.000	0.000
SD	0.937	0.607	1.000	1.000	0.000	0.000

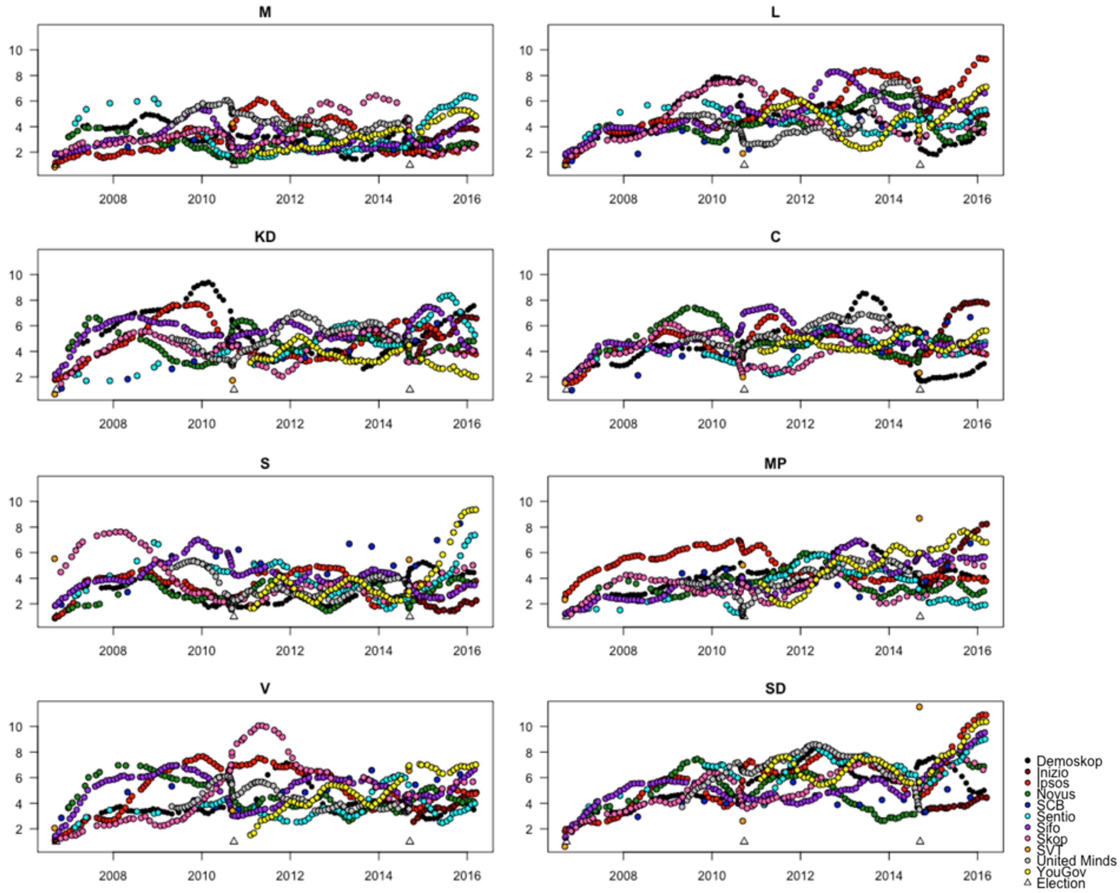

Figure B.9.: Time series of expected value of posterior distribution for the house effects for the different polls.

Table B.8.: Expected value of the posterior distribution for house effects using time variant house effects.

Time variant multilicative house effects on variance								
	M	L	KD	C	S	MP	V	SD
Demoskop	3.03	4.58	5.37	4.28	3.04	4.38	3.93	5.66
Inizio	0.75	2.10	1.65	1.54	2.22	2.11	1.91	0.47
Ipsos	3.16	5.82	4.83	4.45	3.40	4.91	5.42	6.09
Novus	2.71	4.37	4.64	4.77	2.79	3.58	4.61	5.11
SCB	2.63	3.36	4.11	3.92	5.05	3.74	4.86	4.09
Sentio	3.36	4.61	4.41	4.12	4.14	3.23	3.76	6.26
Sifo	3.13	5.39	5.61	5.09	4.21	4.19	5.08	5.30
Skop	3.56	4.68	4.11	4.26	3.64	3.29	5.39	5.27
SVT	1.81	1.70	1.24	0.59	2.52	4.08	3.13	3.73
United Minds	4.35	4.48	4.74	5.10	3.34	3.56	4.10	5.08
YouGov	3.03	4.58	5.37	4.28	3.04	4.38	3.93	5.66

B.5. Plots for simulation study of techniques for handling periodically collected data if latent variable is generated from Dirichlet distribution in section 4.1.2

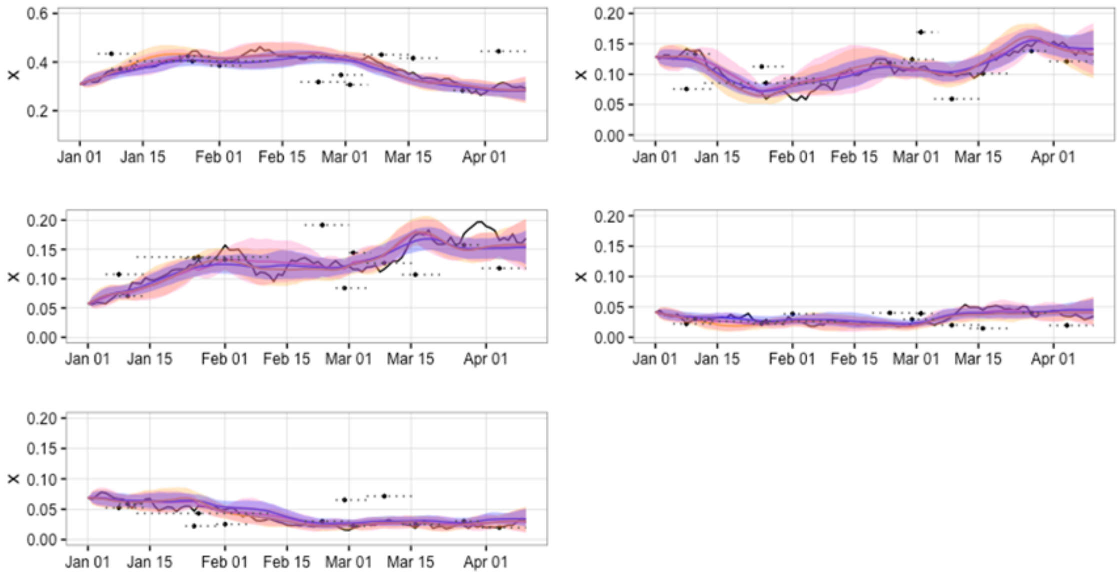

Figure B.10.: Known values of the "latent" variable with the expected value of the posterior distributions for latent variable for the 5 other “parties” using data pre-processing technique 1 (pink), 2 (blue) and 3 (orange). Points are the dates polls are entered into the model using data pre-processing technique 1, dotted horizontal lines are the whole collection period, bands are 95% credible bands.

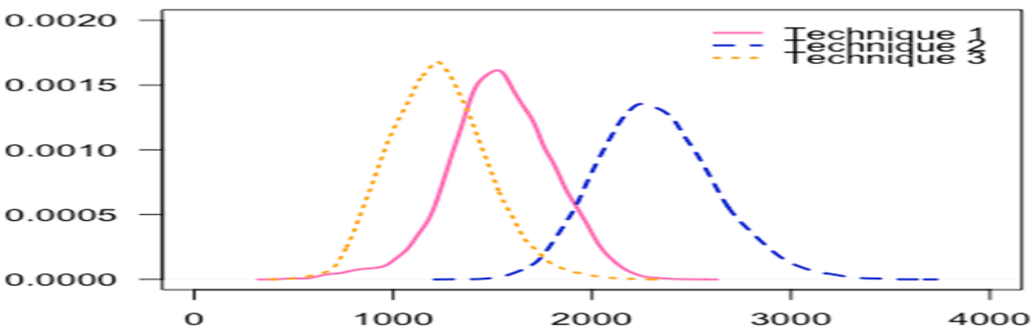

Figure B.11.: Posterior distributions of the concentration parameters in the simulation study for the different data pre-processing techniques.

B.6. Plots and tables for results in section 4.6 – Basic Dirichlet-Dirichlet model

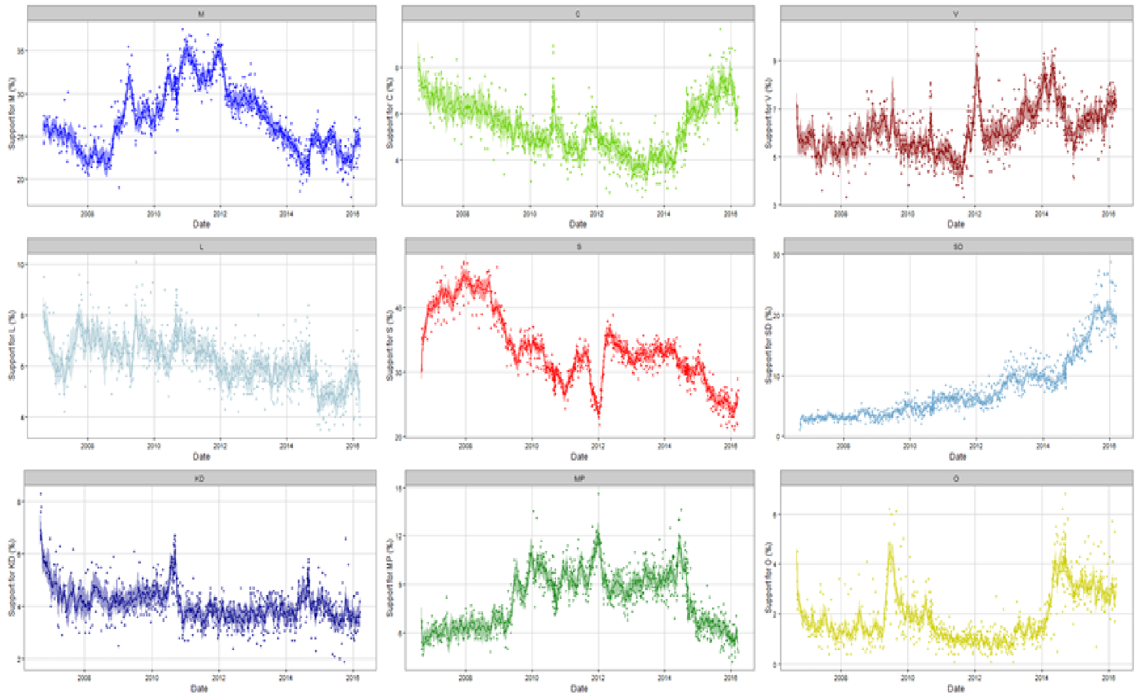

Figure B.12.: Expected value of the posterior distributions for the latent variable for all political party using a Dirichlet-Dirichlet model, with 95% credible bands.

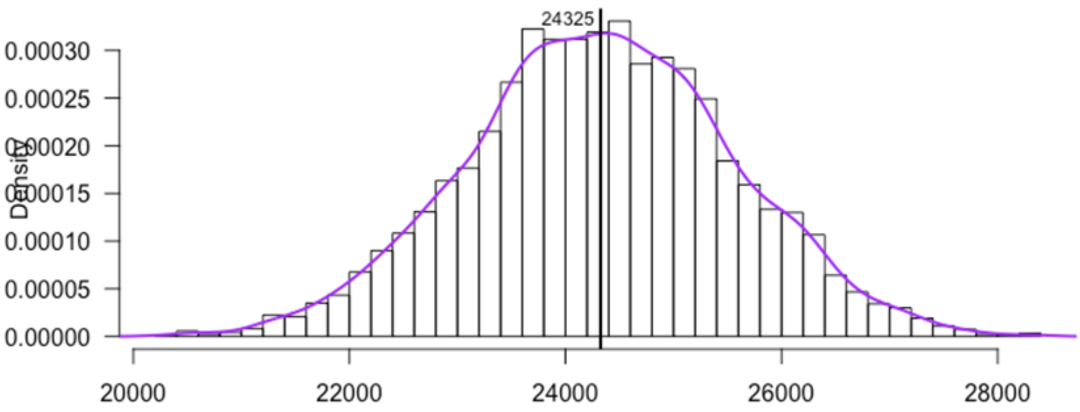

Figure B.13.: Posterior density of the concentration parameter in the basic Dirichlet-Dirichlet model

Table B.9.: The probability of the test quantities minimum, maximum, mean and variance in the replicated data are greater than in the observed data using basic Dirichlet-Dirichlet model. The percentage of negative values and vote intention proportion larger than 1 to the left.

	Basic Dirichlet-Dirichlet					
	Min	Max	Mean	Var	Negative values	Above one
M	0.986	0.453	0.609	0.049	0.000	0.000
L	0.468	0.087	0.126	0.014	0.000	0.000
KD	0.159	0.001	0.000	0.000	0.000	0.000
C	0.997	0.997	1.000	1.000	0.000	0.000
S	0.874	0.444	1.000	0.342	0.000	0.000
MP	0.999	1.000	1.000	1.000	0.000	0.000
V	0.000	0.000	0.000	0.000	0.000	0.000
SD	0.997	0.000	0.000	0.000	0.000	0.000
O	0.997	0.003	0.001	0.000	0.000	0.000

B.7. Plots and tables for results in section 4.7 – Dirichlet-Dirichlet model with a time variant concentration parameter

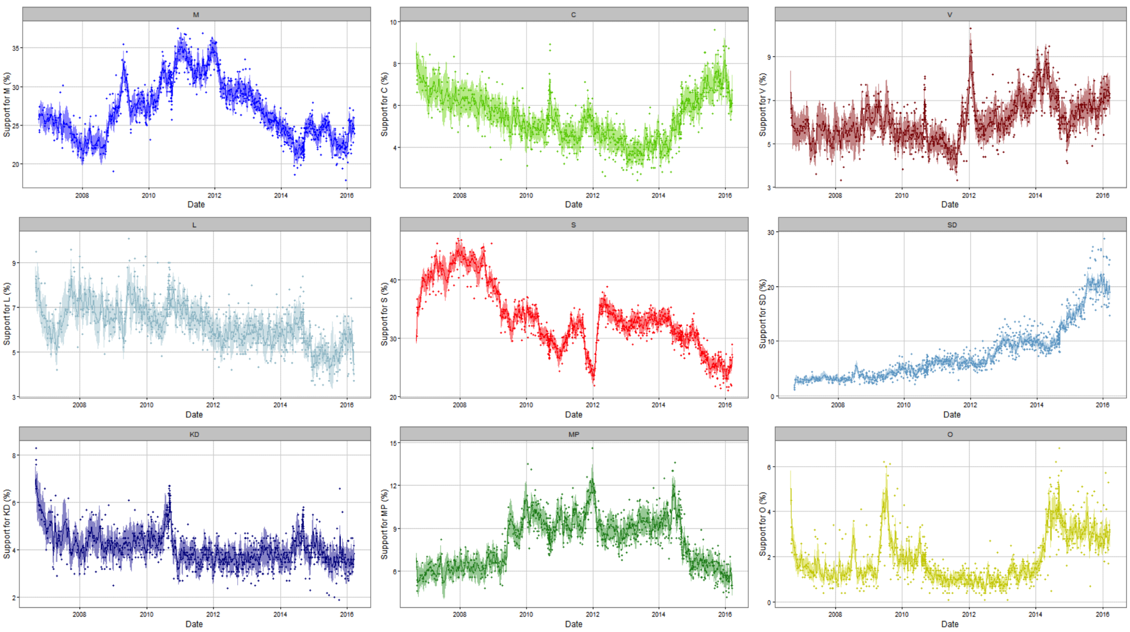

Figure B.14.: Expected value of the posterior distributions for the latent variable for all political party using a Dirichlet-Dirichlet model using a dynamic concentration parameter, with 95% credible bands.

Table B.10.: The probability of the test quantities minimum, maximum, mean and variance in the replicated data are greater than in the observed data using Dirichlet-Dirichlet model with time variant concentration parameter . The percentage of negative values and vote intention proportion larger than 1 to the left.

	Dirichlet-Dirichlet with time variant concentration parameter					
	Min	Max	Mean	Var	Negative values	Above one
M	0.969	0.472	0.605	0.097	0.000	0.000
L	0.295	0.149	0.107	0.098	0.000	0.000
KD	0.090	0.000	0.000	0.000	0.000	0.000
C	0.993	0.998	1.000	1.000	0.000	0.000
S	0.778	0.460	1.000	0.506	0.000	0.000
MP	0.997	1.000	1.000	1.000	0.000	0.000
V	0.000	0.000	0.000	0.000	0.000	0.000
SD	0.986	0.000	0.000	0.000	0.000	0.000
O	0.997	0.006	0.002	0.000	0.000	0.000

C. Notation

M: the Moderates, seen as a right-wing party by voters and part of the Alliance.

L: the Liberals, seen as a right-wing party by voters and part of the Alliance.

KD: the Christian Democrats, seen as a right-wing party by voters and part of the Alliance.

C: the Centre Party, seen as a right-wing party by voters and part of the Alliance.

S: the Social Democrats, seen as a left-wing party by voters.

MP: the Green Party, seen as a left-wing party by voters.

V: the Left Party, seen as a left-wing party by voters.

SD: the Swedish Democrats, seen as a right-wing party by voters.

O: Other parties

Bibliography

Bergman, J. (2015). Are there house effects in Swedish polls? A compositional appraoach. Working Papers in Statistics.

Bergman, J. and Holmquist, B. (2014). Poll of polls: A compositional loess model. *Scandinavian Journal of Statistics*, 41:301 – 310.

Christensen, L. (2015). Påverkar mätmetoden SD:s opinionsstöd? Redux. `http://lovechristensen.github.io/`. Blog; accessed: 2016-04-02.

Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford Statistical Science Series. Oxford University Press, 2nd edition.

Eady, G. (2015). Correcting for (Partisan) Bias in Election Campaign Polls. Working Papers in Statistics.

Fisher, S., Ford, R., Jennings, W., Pickup, M., and Wlezien, C. (2011). From polls to votes to seats: Forecasting the 2010 British general election. *Electoral Studies*, 30:250 – 257.

Frigyik, B. A., Kapila, A., and Gupta, M. R. (2010). Introduction to the dirichlet distribution and related processes. Technical Report UWEETR-2010-0006, Department of Electrical Engineering, University of Washignton.

Gelman, A., Carlin, J. B., Stern, J. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian data analysis*. Taylor and Francis, 3rd edition.

Givens, G. H. and Hoeting, J. A. (2013). *Computational Statistics*. Wiley, 2nd edition.

Jackman, S. (2005). Pooling the polls over an election campaign. *Australian Journal of Political Science*, 40:499 – 517.

Linzer, D. (2013). Dynamic Bayesian forecasting of presidential elections in the states. *Journal of the American Statistical Association*, 108:124 – 134.

Lohr, S. L. (2010). *Sampling: Design and Analysis*. Brooks/Cole, 2nd edition.

Louwerse, T. (2015). Improving Opinion Poll Reporting: The Irish Polling Indicator. Unpublished paper.

Novus (2016). Ekot/Novus Poll of Polls. `http://novus.se/valjaropinionen/ekotnovus-poll-of-polls/`. Online; accessed: 2016-02-04.

Oscarsson, H. (2016). Flytande väljare. Technical Report Democracy Statistics Report 21, Statistics Sweden.

Oscarsson, H. and Holmberg, S. (2011). Swedish voting behaviour. Technical Report Report 2011:4, Swedish National Election Studies Program, Department of Political Science, University of Gothenburg.

Oscarsson, H. and Holmberg, S. (2015). Swedish voting behaviour. Technical Report Report 2015:05, Swedish National Election Studies Program, Department of Political Science, University of Gothenburg.

Pasek, J. (2015). The Polls - Review Predicting Elections: Considering Tools to Pool the Polls. *Public Opinion Quarterly*, 72:594 – 619.

Petris, G., Petrone, S., and Campagnoli, P. (2009). *Dynamic linear models*. Springer New York, 1st edition.

Pickup, M. and Johnston, R. (2007). Campaign trial heats as electoral information: evidence from the 2004 and 2006 Canadian federal elections. *Electoral Studies*, 26:460 – 476.

Särkää, S. (2013). *Bayesian filtering and smoothing*. IMS Textbooks series. Cambridge University Press, 3rd edition.

Shirani-Meh, H., Goel, S., Rothschild, D., and Gelman, A. (2015). Disentangling total error, bias, and variance in election polls. Under review.

SR (2008). Uppdrag. `http://sverigesradio.se/sida/gruppsida.aspx?programid=3113&grupp=5119&artikel=1971435/`. Online; accessed: 2016-04-18.

Stoltenberg, E. A. (2013). Bayesian Forecasting of Election Results in Multiparty Systems. Master's thesis, Univeristy of Oslo.

Sundell, A. (2015). En ranking av svenska väljarbarometrar . `https://politologerna.wordpress.com/2015/10/16/en-rankning-av-svenska-valjarbarometrar/`. Blog; accessed: 2016-02-24.

SVT (2014). SVT:s vallokalsundersökning Riksdagsvalet 2014. `http://www.svt.se/svts/article2323667.svt/binary/SVT_ValuResultat_riksdagsval_2014_PK_0914.pdf`. Accessed: 2016-02-02.

Valmyndigheten (2016). Tidigare val. `http://www.val.se/tidigare_val/`. Online; accessed: 2016-01-20.

Walther, D. (2015). Picking the winner (s): Forecasting elections in multiparty systems. *Electoral Studies*, 40:1 – 13.

Wang, W., Rothschild, D., Goel, S., and Gelman, A. (2015). Forecasting elections with non-representative pollss. *Electoral Studies*, 31:980 – 991.

LIU-IDA/STAT-A–16/004—SE